Министерство образования и науки Российской Федерации
ФГОУ ВО «Уральский государственный педагогический университет»
Институт математики, информатики и информационных технологий
Кафедра информатики, информационных технологий и
методики обучения информатике

Разработка системы фильтрации контента в образовательных учреждениях в рамках проекта "Selecta"

Выпускная квалификационная работа
бакалавра по направлению подготовки
02.03.02 - Фундаментальная информатика и
информационные технологии

Исполнитель:
студент группы Б-41
Института математики, информатики
и информационных технологий
Менщиков А. Е.
Руководитель:
к.т.н, доц. кафедры. ИИТиМОИ
Емельянов Д. А.

Работа допущена к защите
«____»_____ 2017 г.
Зав. кафедрой _____

Екатеринбург – 2017

Реферат

Менщиков А. Е. РАЗРАБОТКА СИСТЕМЫ ФИЛЬТРАЦИИ КОНТЕНТА В ОБРАЗОВАТЕЛЬНЫХ УЧРЕЖДЕНИЯХ В РАМКАХ ПРОЕКТА "SELECTA" выпускная квалификационная работа: 48 стр., рис. 10, 41 библ. назв., приложений 1.

Научный руководитель: к.т.н, доц. каф. ИИТ и МОИ Емельянов Д. А.

Ключевые слова: КОНТЕНТНАЯ ФИЛЬТРАЦИЯ, СИСТЕМЫ КОНТЕНТНОЙ ФИЛЬТРАЦИИ, СЕРВЕР, АНОНИМНОСТЬ, ЦЕНЗУРА.

Объект разработки - система контентной фильтрации.

Цель работы – разработка системы контентной фильтрации для учебных заведений.

В работе описаны результаты проектирования и программной реализации системы контентной фильтрации.

Первая глава содержит теоретический материал, раскрывающий цели, методы и средства по обеспечению анонимности в сети и фильтрации трафика в Интернете.

Вторая глава включает в себя практический аспект работы, а именно описание разработанной системы контентной фильтрации на основе ОС Debian и прокси-сервера Squid.

Проект «Selecta» был успешно внедрен в сетях образовательных учреждений республики Бурятия и получены положительные отзывы о его работе.

Оглавление

Введение....4
Глава 1. Обзор технических средств и правовых ограничений фильтрации нежелательного контента 6
1. 1. Фильтрация нежелательного контента....6
1.1.1. Цель фильтрации нежелательного контента....6
1.1.2. Законы, регулирующие сеть Интернет....7
1.1.3. Анонимность в сети....9
1.1.4. Технические методы фильтрации сетевого трафика....12
1. 2. Существующие современные контент-фильтры....16
2.2.1. Система «NetPolice Pro»....17
2.2.2. Система «Интернет-цензор»....18
2.2.3. Traffic Inspector для школ....19
2.2.4. SquidGuard....20
1. 3. Формализованное описание технического задания на разработку системы контентной фильтрации....21
Глава 2. Разработка и внедрение системы контентной фильтрации....24
2. 1. Описание внутреннего устройства системы контентной фильтрации....24
2. 2. Описание веб-интерфейса....32
2.2.1. Вкладка «Пользователи»....34
2.2.2. Вкладка «Профили»....35
2.2.3. Вкладка «Группы слов»....37
2.2.4. Вкладка «Брандмауэр»....38
2.2.5. Вкладка «Учетные записи»....39
2.2.6. Вкладка «О продукте»....40
2.2.7. Вкладка «Журнал»....41
2.2.8. Вкладка «Настройки»....42
2. 3. Результаты апробации, техническая документация....42
Заключение....43
Список литературы....44
Приложение....48

Введение

В 2006-2008 годах в рамках реализации приоритетного национального проекта «Образование» было осуществлено подключение более 50 000 образовательных учреждений Российской Федерации к сети Интернет. Однако наряду с полезной и необходимой информацией, способствующей получению новых знаний и построению эффективного процесса обучения, ученики получили доступ к ресурсам, содержащее неэтичное и агрессивное содержание.

С целью защиты учащихся образовательных учреждений РФ от противоправного и агрессивного контента в 2006-2007 годах была разработана и внедрена федеральная Система исключения доступа к Интернет-ресурсам, несовместимых с задачами воспитания и образования обучающихся РФ (СИД). Согласно федерального закона Российской Федерации от 29 декабря 2010 г. № 436-ФЗ «О защите детей от информации, причиняющей вред их здоровью и развитию» в срок до 1 сентября 2012 г. во всех образовательных учреждениях на территории Российской Федерации обязательным требованием является установка персонального контент-фильтра на каждый компьютер.

Контент-фильтр — устройство или программное обеспечение для фильтрации сайтов по их содержимому, не позволяющее получить доступ к определённым сайтам или услугам сети Интернет. Средствами контент-фильтрации (далее СКФ) доступа к сети Интернет являются аппаратно-программные или программные комплексы, обеспечивающие ограничение доступа к Интернет-ресурсам, не совместимым с задачами образования и воспитания обучающихся.

Объект исследования – блокировка нежелательного контента при работе с компьютерными сетями.

Предмет исследования – изучение возможностей системы фильтрации нежелательного контента при сетевом трафике.

Целью данного проекта является разработка системы контентного анализа для защиты учебных заведений от несанкционированного контента в рамках проекта «Selecta». Разработка методических указаний для использования данной системы.

Поставленная цель достигается решением следующих задач.

- Рассмотреть локальные нормативные акты и методические материалы для обеспечения информационной безопасности учащихся при использовании ресурсов сети "Интернет".
- Рассмотреть существующие методы фильтрации трафика.
- Рассмотреть способ организации контентной системы в образовательном учреждении.
- Разработать систему контентного анализа в рамках проекта «Selecta».
- Протестировать данную систему.
- Внедрить данную систему в реальный учебный процесс.
- Составить рекомендации по использованию системы контентной фильтрации «Selecta».

Глава 1. Обзор технических средств и правовых ограничений фильтрации нежелательного контента

1. 1. Фильтрация нежелательного контента

1.1.1. Цель фильтрации нежелательного контента

Интернет — всемирная система объединённых компьютерных сетей для хранения и передачи информации. Интернет появился в 60-х годах XX века на основе сетей ARPANET [1]. Основными концепциями сети Интернет являлись:

- *Децентрализованность* сети – все компьютеры сети являются равноправными.
- *Пакетная передача данных* – данные разбиваются на небольшие куски, каждый из которых отправляется по своему маршруту.

Интернет развивался, однако широкую популярность получил после изобретения Всемирной Паутины в 1991 году ученым Т. Бернерс-Ли. Всемирная паутина (World Wide Web) – компьютерная сеть на основе Интернет, предоставляющая доступ к связанным между собой документам [2]. Чаще всего документами является гипертекст на языке разметки HTML. Изначально для просмотра документов во всемирной паутине предлагался протокол прикладного уровня Gopher, но потом его быстро заменил HTTP [3].

Всемирная паутина задумывалась как децентрализованная сеть, открытая для расширения и добавления новой информации, связывающая все документы воедино через систему уникальных идентификаторов – URI [4]. Любой человек, владеющий знаниями (а часто и без них) HTML, может создать и опубликовать

во Всемирной паутине любую информацию или любой файл. Таким образом, в Интернете может быть опубликована любая информация и любые знания, как положительные (учебная информация, учебные видеоролики, книги, журналы, рисунки и т.д.), так и негативные (порнография, изготовление наркотических и взрывчатых веществ и т.д.).

Интернет – хороший источник знаний для учебного процесса, т.к. предоставляет множество источников, начиная от Википедии до онлайн-библиотек и сканов оригиналов документов. Тем не менее, в сети Интернет также много ресурсов развлекательного, порнографического, экстремистского характера, которые следует отсекать от учебного процесса в школах.

1.1.2. Законы, регулирующие сеть Интернет

Интернет долгое время развивался на принципах саморегулирования. Однако рост популярности и использование всемирной паутины для коммерческих целей потребовало внесение норм права для регулирования правоотношений в сети Интернет.

В различных странах в зависимости от распространенности Интернета действуют свои законы и системы фильтрации.

Северная Корея (КНДР) имеет крайне ограниченный выход в мировой Интернет, однако существует внутренняя изолированная сеть, которая называется Кванмён. Доступ в Интернет имеется только у ограниченного количества учреждений [5]. Кванмён насчитывает порядка 1-1.5 тысяч сайтов, в основном научных и учебных организаций. Доступ к Кванмёну осуществляется по телефонным линиям Dial-Up. Согласно утечки данных в 2016 году, в интернет-сегменте, принадлежащем КНДР, зарегистрировано всего 28 сайтов [6].

Китай. Здесь Интернет доступен для широкого круга людей, однако

ограничен и фильтруется в рамках проекта «Золотой щит» [7]. Всех внутренних веб-сайтов требуется регистрация при создании, а публикация новостей из мирового Интернета запрещена без специального одобрения. Фильтрация страниц происходит на основе «черного списка» сайтов и по ключевым словам в тексте [8]. Для фильтрации между мировым интернетом и локальным китайским используется «Великий Китайский Фаервол». Через Фаервол проходит весь трафик между Китаем и прочими странами. В результате скорость доступа к иностранным сайтам падает на порядок, что подстегивает развитие местных сайтов. В Китае блокируются популярные западные социальные сети, поисковые системы (Yahoo, Google) ограничивают политически мотивированные результаты. В качестве альтернативы предлагается внутренние сервисы. Вместо поиска Google популярен местный поисковик Baidu, вместо Amazon предлагается использовать Taobao.com или AliExpress.com [9].

США. Фильтрация сетевого контента запрещена 1-й поправкой к конституции. Запрет на распространение материалов в сети может быть связан лишь с нарушением положений законодательства о клевете, детской порнографии, интеллектуальной собственности. Кроме того, некоторые школы и библиотеки блокируют на своих компьютерах доступ к вредной для детей информации. Фильтрацию Интернет пытаются ввести в рамках борьбы с пиратством – законопроекты SOPA (2011) и PIPA (2011) [10], но оба отложены [9].

РОССИЯ. В российском сегменте Интернет (Рунет) такими документами являются Федеральный закон № 149 «Об информации, информационных технологиях и о защите информации» от 27.07.2006 [11] и ряд поправок Федеральный закон № 436 «О защите детей от информации, причиняющей вред их здоровью и развитию» от 29.12.10 [12], Федеральный закон № 139 «О внесении изменений в федеральный закон «О защите детей от информации,

причиняющей вред их здоровью и развитию» от 28.07.12 (статьи 6, 55, 119, 140) [13], Федеральный закон №185 "О внесении изменений в отдельные законодательные акты Российской Федерации и признании утратившими силу законодательных актов (отдельных положений законодательных актов) Российской Федерации в связи с принятием федерального закона "Об образовании в Российской Федерации" от 02.07.2013, статья 29 федерального закона №307 "О внесении изменений в Кодекс Российской Федерации об административных правонарушениях и отдельные законодательные акты Российской Федерации и о признании утратившими силу отдельных положений законодательных актов Российской Федерации в связи с уточнением полномочий государственных органов и муниципальных органов в части осуществления государственного контроля (надзора) и муниципального контроля" от 14.10.2014. Контроль над исполнением этих документов осуществляет ведомство Роскомнадзор.

1.1.3. Анонимность в сети

Под анонимностью в сети подразумевается возможность совершать ряд действий без возможности опознать личность пользователя [16]. Люди не любят, когда за их действиями в сети следят. Какие сайты они посещают? Какие комментарии они оставляют? Чем интересуются?

Из-за кажущейся абсолютной анонимности в сети у людей возникает эффект «Растормаживания в Сети». «Растормаживание» может быть, как положительным, так и негативным в зависимости от деятельности человека в сети [17].

Согласно 23 и 24 статей Конституции Российской Федерации человек имеет право на частную жизнь, тайну переписки и запрет о сборе частной информации о частной жизни лица без его ведома, в том числе и в сети [18].

Таким образом, человек имеет право на соблюдение анонимности в сети. Правительства, работодатели и преподаватели учебных заведений нередко ограничивают доступ к информации и ущемляют свободу выражения подчинённых им людей под всевозможными предлогами. Наблюдается тенденция усиления слежки в Сети.

Компьютеры в сети Интернет однозначно определяются их IP-адресами. Сокрыть IP-адрес можно различными средствами для получения анонимности [20]. Каждый из них имеет свои плюсы и минусы, а для большей защищенности следует комбинировать их вместе [19].

Прокси-сервер. Один из самых простых способов стать анонимным в сети – это использование прокси-сервера. Прокси-сервер, в самом простейшем случае, - сервер, принимающий HTTP-запрос от клиента и передающий HTTP-запрос дальше к исходному серверу. Прокси-сервер может использоваться для проксирования HTTP-трафика, так и более сложные случаи: передача HTTPS-трафика, передача по SOCKS протоколу. Для подключения прокси-сервера достаточно указать его адрес в настройках браузера или в настройках системы [21].

Анонимайзеры. В Интернете можно найти много бесплатных *прокси-серверов*, однако платой за бесплатность является то, что трафик пользователя может модифицироваться (например, для вставки своей рекламы), либо сохраняют запросы пользователя. Многие анонимайзеры и им подобные сервисы также являются всего лишь прокси-серверами для HTTPS. Подходят ли они для полной анонимности в сети? Нет, они всего лишь меняют IP-адрес пользователя, даже так можно отследить пользователя. У анонимайзеров URL запросов отличен от URL исходного запроса, и поэтому такие запросы к заблокированному сайту сложнее отследить, но можно заблокировать сами анонимайзеры, т.к. их список известен и пополняется [20].

VPN. Другой популярный способ анонимности в сети является Virtual

Private Network (VPN) – это технология, обеспечивающая защищённую (закрытую от внешнего доступа) связь логической сети, поверх частной или публичной при наличии высокоскоростного интернета. То есть, получаем шифрованную сеть внутри сети. Шифрованные пакеты пользователя инкапсулируются в пакеты IP или Ethernet и пересылаются по внешней сети к удаленному серверу, который принимает и расшифровывает полученные данные, маршрутизирует пакеты уже во внутренней сети [22].

Коммуникации осуществляются по сетям с меньшим или неизвестным уровнем доверия. Для данного способа требуется свой сервер в другой сети (стране). Для обеспечения полной анонимности удаленный сервер должен приобретаться также анонимно. Это возможно сделать при приобретении сервера с помощью одной из электронных валют, к примеру, биткоин (Bitcoin) [20].

Tor. Это специальная сеть, основанная на луковой маршрутизации. Сеть Tor – система, состоящая из нескольких прокси-серверов. Прокси-сервера ретранслируют между собой шифрованные данные. Некоторые из прокси-серверов являются выходными узлами, которые расшифровывают запросы пользователей сети и отправляют их в открытом виде на ресурсы сети. Пользователи соединяются с выходными узлами через несколько промежуточных узлов. Все узлы выбираются случайно, однако, только один раз при создании соединения. Промежуточные сообщения [23].

Несколько раз шифрованные сообщения проходят через несколько луковых маршрутизаторов, каждый из которых убирает свой слой шифрования. Выходной узел получает расшифрованное сообщение и отправляет его в Интернет. Выходных узлов в мире на данный момент существует примерно 900 штук [24]. Для упрощения подключения к сети Tor делают специализированные сборки браузеров, поддерживающих луковую маршрутизация сразу «из коробки» (не требующую дополнительной настройки). Преимуществом Tor

является легкость использования и высокой скорость, при среднем уровне безопасности. Сайты внутри Tor–сети используют псевдо-DNS с доменом верхнего уровня «.onion» [23].

Tor используют журналисты для анонимной публикации своих статей, преступники, активисты разного рода и просто люди, любящие приватность. У преступников есть большой арсенал средств помимо Tor для анонимизации от смены личности до краденых устройств или доступов в сеть, от ботнетов до вирусов-троянцев.

I2P. Специальная сеть, основанная на тех же идеях, что и Tor. I2P (аббревиатура от invisible internet project проект - «Невидимый интернет») – сеть с шифрованием поверх Интернета. В I2P существуют многие альтернативные реализации сервисов, доступных в Интернет с упором на анонимность. Из I2P можно достигнуть анонимности в веб за счет прокси серверов I2P <-> Веб, однако скорость соединения получается меньше, чем у Tor [25,26].

Для сохранения анонимности использования одних лишь технических средств недостаточно. Часто пользователи сети сами невзначай “деанонимизируют” себя, не смотря на технические средства, своим поведением, оставшимися деталями. К примеру, если пользователь не удалит метаданные из фотографии и выложит их в сеть, то из этих данных можно найти геотеги или автора фото.

Использование одной из технологии даёт базовую анонимность, достаточную для простого обхода фильтров. Однако для серьезной безопасности и анонимности желательно использовать одновременно несколько средств [19].

1.1.4. Технические методы фильтрации сетевого трафика

Существует немало способов фильтрации трафика на различных уровнях

TCP/IP стека. Каждый TCP/IP-пакет характеризуется 4 параметрами: IP адрес источника, IP-адрес назначения, порты источника и назначения. Зная IP-адрес источника можно определить кто из пользователей послал этот пакет, а зная IP-адрес и порт назначения можно понять, кому предназначен этот пакет и необходимо ли проверять данный пакет и всю TCP/IP сессию. Собирая и проксируя трафик нужных TCP/IP сессий получаем дополнительные сведения для фильтрации HTTP(S) запросов, такие как URL запроса, домен и тело запроса. Для полученных сведений можно использовать различные методы фильтрации.

Блокирование по IP-адресу. При применении данного метода сервер, на котором находится нежелательный материал, становится полностью недоступным для пользователя. Главным преимуществом этого метода является его простота – он может быть реализован с помощью базового сетевого оборудования, используемого интернет-провайдерами. Однако с учетом современных технологий по одному IP-адресу могут находиться тысячи сайтов, а также других сервисов, таких как FTP или электронная почта, поэтому его блокирование приведет к тому, что все они станут недоступны. Из-за низкой точности данного метода страны применяют его с осторожностью. Блокирование по IP-адресу легко обходится при помощи различных технических решений, в частности, прокси-серверов и VPN.

Искажение DNS-записей. При обращении пользователя к любому сайту, компьютер посылает запрос к DNS-серверу для того, чтобы преобразовать доменное имя в IP-адрес. В случае применения данного метода, DNS сервер возвращает неверный адрес, и сайт оказывается недоступным. Искажение DNS-записи также может быть реализована без применения дополнительного оборудования. Ее преимуществом перед блокированием по IP-адресу является более высокая точность – недоступным становится только один сайт на сервере. При этом все равно происходит чрезмерное блокирование. Например,

Китай периодически лишает своих пользователей доступа к CNN International из-за появляющихся там нежелательных новостей. Хотя ставится цель фильтрации только одной страницы новости, остальные страницы сайта также становятся недоступны. Искажение DNS-записей легко обходится пользователями – в настройках операционной системы достаточно указать альтернативный DNS-сервер или вручную прописать IP-адрес заблокированного сайта.

Блокирование по URL-адресу. В HTTP-протоколе URL-адрес содержит доменное имя сайта, а также параметры запроса. Они могут быть сверены со списком заблокированных ключевых слов, и в случае соответствия, связь пользователя с запрошенным ресурсом разрывается, или он перенаправляется на блок-страницу. Данный метод является более эффективным по сравнению с блокированием по IP-адресу и искажением DNS-записи, но требует дополнительного оборудования, так как использует поверхностный анализ пакетов. Его дополнительным преимуществом является то, что он способен динамически блокировать новые страницы, если в их адресе содержатся запрещенные слова. Например, в Китае блокируются все запросы, содержащие слова “falun” и “gong”. Однако при неправильной настройке ключевых слов точность метода резко ухудшается – он может пропускать нежелательный материал или, наоборот, допускать чрезмерное блокирование. Блокирование по URL-адресу нельзя обойти с помощью обычных прокси-серверов – необходимы инструменты, которые шифруют трафик, такие как VPN или TOR

Блокирование по типу файла. В ответе на HTTP запрос сервер устанавливает заголовок *Content-Type*, в котором описывается тип передаваемого. Значением этого заголовка является один из подходящих MIME-тип. MIME – стандарт передачи различных типов данных по электронной почте, а также спецификация для кодирования информации и форматирования сообщения. Сопоставляя значения заголовка Content-Type и

запрещенные для пересылки типы файлов можно фильтровать запросы. Этот метод не является надежным средством для блокировки.

Фильтрация HTTPS-запросов. Заголовки и тело HTTP-запросов передаются в открытом виде, поэтому фильтр может их выделить и использовать для проверки сайта. Однако для страниц HTTPS сайтов невозможно узнать заголовки запросов из-за шифрования трафика. Поэтому для HTTPS-запросов фильтры производят атаку MiTM [40] и подменяют полностью или частично все сертификаты сайтов [28, 29, 30]. Важным расширением протокола TLS является SNI – Server Name Indication. С помощью этого расширения домен доступен в открытом виде. Это расширение используется для организации нескольких HTTPS-сайтов на одном IP-адресе, но также дает возможность фильтру подменять сертификаты только для определенных сайтов [41].

Пакетная фильтрация. Наиболее сложный и дорогостоящий метод, так как он требует применения глубокого анализа пакетов. На данный момент полноценно реализован только в Китае. При использовании пакетной фильтрации, изучаются не только заголовки пакетов, содержащих URL адрес, но и все их содержимое. В случае наличия запрещенных слов, связь между пользователем и сервером разрывается. Метод позволяет фильтровать нежелательный контент не только в веб-страницах, но и во всех протоколах – электронной почте, сервисах мгновенных сообщений и др. Существенным недостатком данного метода является то, что применение глубокого анализа пакетов может привести к снижению скорости интернет-соединения, что, к примеру, наблюдается при доступе из Китая к зарубежным интернет-серверам. В остальном, пакетная фильтрация обладает теми же достоинствами и недостатками, что и блокирование по URL-адресу.

Фильтрация через HTTP прокси-сервер. Данный метод наиболее часто используется организациями для подключения корпоративных сетей к

Интернету, но его можно использовать для фильтрации интернета в рамках всей страны. Гибридный вариант под названием Cleanfeed [39] эффективно применяется в Великобритании и Канаде для борьбы с детской порнографией. Каждый запрос пользователя сверяется со списком IP-адресов, содержащих запрещенные материалы. Если совпадений нет, то запрос пользователя отсылается напрямую. В противном случае, он перенаправляется на прокси-сервер общественной организации Internet Watch Foundation. Прокси-сервер получает запрашиваемую страницу и анализирует ее на наличие детской порнографии. Если страница не содержит запрещенных материалов, то пользователь получает к ней доступ, иначе – создается видимость, что ресурс недоступен. Гибридные варианты фильтрации через HTTP прокси-сервер позволяют при низкой стоимости точно блокировать узкие категории контента. При этом, они столь же легко обходятся, как и фильтрация по IP-адресу.

Фильтрация результатов поиска. В ряде стран, таких как Китай, Франция и Германия, работающие там поисковые системы обязаны исключать из результатов поиска ссылки на запрещенные материалы. Так, во французских и германских версиях Google из поисковых результатов исключаются ссылки на неонацистские группы и другие материалы, запрещенные законом. Таким образом, пользователи не могут найти нежелательный контент. Фильтрация результатов поиска – один из основных методов борьбы с нарушениями авторских прав в Интернете. Метод обходится использованием других поисковым систем – например, международная версия Google не исключает из результатов сайты неонацистских группировок и при этом доступна из Франции и Германии [27].

1. 2. Существующие современные контент-фильтры

Все фильтры делятся на две категории:

- личные, т.е. устанавливаются на компьютер пользователя;
- виде сервера (маршрутизатора), где сетевой трафик от всех пользователей собирается различными методами и фильтруется.

Каждый из способов имеет свои преимущества и недостатки. При установке фильтра на компьютер пользователя не нужен отдельный сервер для фильтрации, но это преимущество нивелируется необходимостью установки и настройки каждого экземпляра приложения отдельно, а также сложностью сбора статистики по запросам пользователей. При установке фильтра на сервер, администратор получает простой и единый способ для настройки и управления фильтрацией. В зависимости от сложности требуемых настроек, и количества хостов в сети администратором выбирается оптимальный способ организации фильтрации в образовательных учреждениях. Для небольшого количества хостов в сети и малых знаний администратора данной сети оптимально установить личные контент-фильтры на каждый хост в сети, а для средних и крупных сетей необходимо использовать вариант с выделенным сервером для фильтрации трафика.

2.2.1. Система «NetPolice Pro»

Персональный контентный фильтр NetPolice Pro позволяет установить необходимый уровень доступа к ресурсам сети Интернет в школах, обеспечивая эффективную фильтрацию контента по спискам категорий, рекомендованным Минобрнауки России. При использовании NetPolice Pro обеспечивает высокий уровень безопасности за счет реализации в решении двух технологий: фильтрации URL и динамической фильтрации. Фильтр проверяет к какой категории (запрещенной или разрешенной) относится запрашиваемый сайт, но и анализирует его содержимое веб-страниц. Программа NetPolice Pro позволяет редактировать списки слов, которые будут блокироваться непосредственно в

поисковых запросах, а также вести «черный» и «белый» списки ресурсов. Фильтр настраивается только по предварительно установленному паролю, и, даже обладая администраторскими правами на компьютере, пользователь не сможет отключить защиту NetPolice Pro. Поддерживается 9 категорий с потенциально опасным содержанием, по 7 категорий для сайтов несовместимых с задачами образования и ресурсов с неконтролируемым содержанием и 50 профессиональных категорий [31,32].

Достоинства:

- не требует отдельного мощного сервера для анализа,
- находится в Едином реестре российских программ для электронных вычислительных машин и баз данных.

Недостатки:

- фильтрует браузеры и мессенджеры,
- фильтрация происходит только по протоколу HTTP (отсутствует поддержка HTTPS)
- предназначен только операционные системы семейства Windows (хотя есть отдельная версия под некоторые дистрибутивы Linux).

2.2.2. Система «Интернет-цензор»

Бесплатный интернет-фильтр для детей Интернет цензор под ОС Windows. Программа ставится на личный компьютер пользователя администратором, единого интерфейса настройки нескольких пользователей нет. Фильтрация происходит по принципу "белых списков". База «белых списков» сайтов включает в себя порядка миллиона вручную проверенных сайтов из безопасных сайтов Рунета и основных иностранных ресурсов. Программа защищена от взлома и обхода фильтрации паролем. На 2016 год можно найти основной сайт программы не работает, однако саму программу

можно найти в сети Интернет [33,34].

Достоинства:

- бесплатный,
- не требует отдельного мощного сервера для анализа,
- простая установка.

Недостатки:

- анализ трафика основан на списке вручную проверенных сайтов,
- фильтрует браузеры и мессенджеры,
- поддерживается только ОС семейства Windows
- отсутствует фильтрация по категориям
- низкое качество фильтрации.

2.2.3. Traffic Inspector для школ

Система от российской компании Smart-Soft – Traffic Inspector – устанавливается на шлюз школы. Программа рассчитана на использование в среде операционных систем Microsoft Windows 7, Windows 8, Windows 8.1, Windows Server 2008 R2 x64, Windows Server 2012, Windows Server 2012 R2. Начиная с версии 3.0.1 (2013 год) поддерживается не только 32-х битные системы, но и 64-битные. Traffic Inspector имеет лицензию и сертификат. ФСТЭК по 3 классу защищенности. Поддерживает фильтрацию по URL («черные» и «белые» списки), по категориям. Ведется статистика, учет и биллинг (расчет) трафика пользователей и статистика посещений. Поддерживаются различные виды авторизации пользователя, в том числе интеграция с AD. В дополнение к основному продукту имеются модули расширения для фильтрации рекламы, фишинга, проверка трафика через антивирус, контентная фильтрация. Администрирование программы осуществляется в графическом режиме, через оснастку Microsoft Management

Console. Минусом является долгая разработка проекта, а также, что решение на основано на ОС Windows и использует сетевой драйвер в режиме ядра из-за чего при неполадках драйвера вся система может упасть с потерей данных [35,36].

Достоинства:

- собирается статистика по запросам;
- находится в едином реестре программного обеспечения;
- имеет аппаратное решение;
- ведется статистика запросов;
- имеется множество вариантов аутентификации пользователей.

Недостатки:

- как правило, Windows не слишком хорошо подходит для создания файерволла;
- для настройки требуются знания по администрированию Windows систем.

2.2.4. SquidGuard

SquidGuard – open-source программа, дополняющая прокси-сервер Squid. В файл с настройками Squid [37] администратор добавляет вызов сторонней программы – редиректора URL. SquidGuard может фильтровать сайты по доменам, по определенному времени суток и IP пользователя или IP цели. Заблокированные страницы логируются. Минусы: отсутствует контентная фильтрация, настройка требует от администратора кроме знаний по настройке Squid также знаний синтаксиса правил по настройке SquidGuard, версия 1.5 2010 года является последней [38].

Достоинства:

- бесплатная,

- имеется логирование пользовательских запросов.

Недостатки:

- уже не поддерживается,
- сложная в настройке
- требует знаний по администрированию прокси-сервера Squid и Linux-систем.

1. 3. Формализованное описание технического задания на разработку системы контентной фильтрации

1. Введение, общие сведения о создаваемой системе.

Название: Ideco «Selecta».

Область использования: ВУЗ, предприятие, щкола.

Данные об авторе: команда разработки проекта Ideco Selecta, в состав которой также входит студент группы Б-41, Института математики, информатики и информационных технологий, Менщиков А. Е.

Руководитель разработки: ведущий программист IDECO SELECTA Владимиров Антон Павлович.

Основания и назначение разработки: объектом разработки является система контентной фильтрации, с простым понятным интерфейсом и возможностью индивидуальной политики фильтрации для разных сетей и IP-адресов.

Место внедрения и заказчики: бурятский филиал «Ростелеком» (ОАО «Ростелеком») для внедрения в общеобразовательных учреждения Республики Бурятия и абонентов, обучающихся в Республиканском центре психолого-медико-социального сопровождения.

2. Требования к продукту разработки.

2.1. Процессы и структуры, в которых предполагается использование продукта разработки.

Продукт разработки предполагается использовать в учебном процессе в качестве средства, защищающего учеников от нежелательной информации во время учебного процесса.

2.2. Характеристика персонала (количество, квалификация, степень готовности)

Для конфигурирования, управления системой контентной фильтрации «Selecta» и осуществления ее технической поддержки не требуются специальные знания и навыки, помимо базовых знаний сетевых технологий.

2.3. Перечень требований к аппаратно-программному окружению.

2.3.1. Локальный сервер, поддерживающий процессор x86-64 и имеющий не менее 8 гигабайт оперативной памяти с как минимум двумя сетевыми картами.

2.3.2. Выход в Интернет.

2.4. Указание программного обеспечения, используемого для реализации.

2.4.1. Локальный компьютер с операционной системой Ubuntu/Debian.

2.4.2. Браузер Mozilla Firefox.

2.4.3. Выход в интернет.

2.4.4. Виртуальная машина KVM, VMware ESX.

2.5. Требования к интерфейсу пользователя: интуитивно понятный, без необходимости обучения администратора

3. Состав и содержание работ по созданию веб-интерфейса.

3.1. Анализ требований к системе.

3.1.1. Вход в веб-интерфейс под уникальным логином и паролем.

3.1.2. Заполнение информации о своем аккаунте.

3.1.3. Возможность создания, редактирования и удаления настроек фильтрации, в частности для пользователей и настроек фильтрации.

3.1.4. Просмотр статистики по пользователю, запросов пользователей.

3.1.5. Возможность поиска по записям студентов.

3.1.6. Наличие интуитивно-понятного интерфейса пользователя.

3.1.7. Фильтрация запросов пользователей по спискам РКН

3.1.8. Фильтрация запросов пользователей при помощи морфологического анализа.

3.2. Проектирование и разработка системы.

3.2.1. Разработка технического задания.

3.2.2. Разработка структуры системы.

3.2.3. Разработка интерфейса.

3.2.4. Реализация системы.

3.2.5. Написание документации.

3.3. Тестирование.

4. Требования к документированию.

4.1. Перечень сопроводительной документации.

- Техническое задание.
- Руководство пользователя.

4.2. Требования к содержанию отдельных документов.

Глава 2. Разработка и внедрение системы контентной фильтрации

2. 1. Описание внутреннего устройства системы контентной фильтрации

Система контентной фильтрации “Selecta” предназначена для фильтрации интернет-трафика для образовательных учреждений и Интернет-провайдеров с категоризацией интернет-трафика по URL и глубокой фильтрацией по содержимому контента.

Программный продукт «Selecta» основан на операционной системы Debian 8. Debian 8 – дистрибутив Linux, являющийся стабильной платформой для создания программ и предоставляющий большое предоставляет множество системных утилит и систем. Например, Systemd – системный демон инициализации других демонов, iptables – утилита для управления правилами фильтрации пакетов.

«Selecta» представляет собой набор небольших демонов – микросервисов. Каждый из сервисов отвечает за свою часть системы. Каждый сервис хранит свои настройки и данные в своей базе данных на диске. Сервисы общаются между собой посредством посылки отдельных небольших сообщений. Сервис формирует сообщение и асинхронно посылает его по протоколу AMQP (версии 0-9-1) отдельного демона - менеджера очередей RabbitMQ. Менеджер очередей RabbitMQ получает сообщения и отправляет в одну или несколько очередей. Сервисы регистрируют в RabbitMQ очереди из которых будут получать сообщения. На каждое сообщение, полученное из RabbitMQ посылает сигнал ACK, означающее, что сообщение получено и обработано. Сообщения для сервиса копятся в очереди, пока предыдущие сообщения не будут отработаны.

Использование отдельного сервиса вместо прямого обращения от сервиса к сервису имеет ряд преимуществ.

- Расширяемость. Несколько экземпляров сервиса могут читать сообщения из одной очереди и таким образом обрабатывать их быстрее.
- Вместо множества соединений с разными сервисами поддерживается только одно с RabbitMQ.
- Количество связей между сервисами растет вообще говоря квадратично с количеством сервисов и их экземпляров. RabbitMQ упрощает взаимодействие между сервисами.

Сервисы написаны на языках программирования Python и JavaScript на программной платформе Node. Наличие сервисов, написанных на разных языках программирования является стандартной практикой для микросервисной архитектур. Под конкретную часть системы – сервис – можно выбрать более подходящий язык. Выбор языков продиктован легкостью и быстротой разработки на данных языках.

Node основан на JavaScript-движке V8. Движок V8 исполняет JavaScript код, ускоряя его за счет компиляции в бинарный код и применения различных эвристик. JavaScript – популярный язык и на данный момент в реестре пакетов **npm** находится примерно 475 000 пакетов.

На Python имеется большое количество библиотек для работы с сетевыми протоколами, легкий синтаксис для написания, дружелюбное комьюнити. Python хорошо подходит для написания серверов и скриптов, однако CPython (основная реализация Python) является интерпретатором и проигрывает в скорости Node.

В качестве хранилища настроек используются БД SQLite. База данных SQLite является встроенной и все данные хранятся в единственном файле на диске. К базе данных возможен доступ на чтение для нескольких программ, однако запись данных может осуществлять только одна. Преимущества SQLite по сравнению с другими СУБД.

- Легкость использования и настройки.
- По сравнению с NoSQL СУБД (такие как Mongo DB) присутствует схема, транзакции и внешние ключи.
- Изолированность хранилищ разных сервисов. Данные одного сервис хранятся в одном месте (файле), когда как при использовании таких СУБД как MongoDB, PostgreSQL etc. Данные будут хранится в фактически в одной базе.
- В SQLite хранятся все настройки пользователей и обычно данных оказывается мало, а их обработка не требует много времени.

Также в качестве отдельного хранилища для статистики посещенных страниц пользователей используется MongoDB. MongoDB не имеет определенной схемы хранящихся данных, не поддерживает сложные транзакции, но зато хорошо подходит для хранения денормализованных данных и работой с логами, текстами и сложными объектами.

Все сервисы «Selecta» являются юнитами systemd. Systemd запускает и следит за состоянием сервисов. Если сервис падает, то systemd перезапускает упавший сервис, а также запускает нужные сервисы по таймеру (более умный аналог демона crond). В юнит-файле сервиса описаны зависимости сервиса от других сервисов и ресурсов (сети, файлов, сокетов, etc.), порядок старта сервисов, аргументы сервисов и другая метаинформация о сервисе. Systemd из

зависимостей строит дерево и сортирует в топологическом порядке, а затем загружает и запускает все нужные сервисы параллельно.

При первоначальной загрузке создаются базы данных, запускаются сервисы и появляется консоль. Администратор настраивает через nmcli интерфейсы сети. Дальнейшая настройка проводится через веб-интерфейс по указанному пользователем IP адресу.

Все запросы к веб-интерфейсе попадают в nginx. Nginx раздает статические файлы, а остальные запросы передает в сервис веб-бэкенда – cf-web-backend. Веб-бэкенд принимает HTTP-запросы, проверяет их правильность, права пользователя и делает запросы к остальным сервисам. Сам веб-бэкенд хранит и обрабатывает всё что связано с аккаунтами веб-интерфейса – права доступа, каких пользователей системы можно редактировать, имена и т.д. Веб-бэкенд не обрабатывает запросы, а передает их дальше по системе RPC нужному сервису. К примеру, все запросы, связанные с AD, поступают в сервис интеграции с AD – ad-backend; запросы, связанные с пользователями системы, профилями и настройками фильтрации поступают в мастер-сервис контентного фильтра – cf-analyzer. Есть также сервисы настроек файерволла (firewalld), сервис лицензирования (licensed), сервис статистики (cf-journal и cf-journal-collectd), сервис настроек прокси-серверов Squid (prepare-proxy), сервис captive portal.

Фильтрация возможна в нескольких режимах: wccp, роутер и в режиме «в разрыв». В режиме wccp весь трафик идет от маршрутизаторов Cisco по протоколу wccp. В режиме роутера «Selecta» настраивается роутер между различными сетями. В режиме работы «в разрыв» внешние и внутренние интерфейсы находятся в одной логической подсети, но разных физических подсетях.

От режима работы зависят настройки файрволла и настройки сетевых интерфейсов. Для настройки межсетевого экрана Netfilter, встроенного в ядро Linux, используются утилиты iptables, ipset и ebtables для управления и фильтрации на сетевом и канальном уровне TCP/IP стека. Все транзитные IP-пакеты, которые предназначены на порты 80 и 443 перенаправляются в прокси-сервер Squid. Порты 80 и 443 являются портами по умолчанию для протоколов HTTP и HTTPS соответственно.

Squid – популярный многофункциональный кэширующий прокси-сервер, написанный на C++. Squid – open-source проект. Множество людей и фирм поддерживают, разрабатывают и используют Squid. Squid имеет поддержку протокола ICAP для адаптации контента, переданного по HTTP(S), поддержку WCCP и различных способов авторизации (к примеру, AD или через Basic Authorization). Важной возможностью Squid является возможность использовать хелперы – дополнительные программы, указывающие подходит данный запрос под ACL. ACL - Access Control List – список контроля доступа, который определяет доступ к ресурсу или ряд действий, происходящих с запросом.

ICAP – Internet Content Adaptation Protocol – протокол для расширения возможностей прокси-сервера. ICAP был представлен в 2003 в RFC 3507 J. Elson и A. Cerpa. ICAP представляет собой легковесный HTTP-подобный протокол для инкапсуляции в себе HTTP трафика и передачи его между прокси-сервером и сторонним адаптирующим сервисом.

Для HTTPS запросов выполняется проверка, на необходимость подмены сертификатов. При включенной авторизации через Адаптивный портал (Captive Portal) выполняется проверка, что такой пользователь уже зарегистрировался и при необходимости выполняется переход на страничку авторизации. А при

интеграции с AD проверяется, что возможность доступа к ресурсу конкретного пользователя.

После выполнения всех проверок для HTTP(S)-запроса на ACL Squid отправляет запрос по протоколу ICAP на сервис контентной фильтрации – cf-analyzer. Из запроса извлекается URL сайта, проверяется на белые и черные списки, категорируется и в зависимости от этого весь запрос либо разрешается (возможно с модификацией), либо запрещается и вместо этого посылается страница блокировки. Если присутствует интеграция с другими ICAP-серверами, то Squid проходит по всей цепочке ICAP-серверов прежде чем послать HTTP(S)-запрос на исходный сервер (origin server).

Исходный сервер выполняет запрос пользователя и отдает некоторый контент – HTML-страничку, картинку, JSON, и т.д. Squid отправляет ответ (response) от исходного сервера в новом ICAP-запросе в cf-analyzer. Сервис фильтрации в зависимости от настроек профиля пользователя проверяет тип файла, а для HTML и текстового контента проверяет его на присутствие запрещенных администратором слов.

Основные сервисы «Selecta»:

- *Content_filter_backend* (разработан в данном проекте) – основной веб-бэкенд. Все HTTP-запросы проксируются через прокси-сервер Nginx и попадают в этот сервис. HTTP-запросы на получение статические файлов (статические HTML-страницы, CSS-файлы, Javascript код веб-интерфейса и прочее) обслуживает Nginx непосредственно. API между веб-интерфейсом и веб-бэкендом выполнено в REST-стиле.
- *Content_filter_analyzer* (разработан в данном проекте) – основной сервис, анализирующий запросы пользователей и ответы веб-серверов.

- *Content_filter_journal* (разработан в данном проекте) – сервис статистики и журнал посещенных ресурсов. Сюда поступает информация о запросах пользователей от content_filter_analyzer. Все запросы агрегируются и записываются в базу данных в журнал.
- *MongoDB* – не реляционная база данных, используемая для хранения настроек и записей журнала. MongoDB выбрана из-за гибкой схемы данных.
- *Cf_monitor* (разработан в данном проекте) – сервис мониторинга состояния сервера. Предоставляет администратору основную информацию по состоянию компьютера:
    - Количество используемой памяти
    - Процент загрузки CPU
    - Нагрузка на сеть и HTTP(S) трафик.
- *Squid* – прокси-сервер

Весь процесс фильтрации можно разбить на несколько этапов:

1. «Selecta» устанавливается на сервер пользователя в «разрыв» сети. Администратор настраивает прозрачный сетевой мост между различными сегментами подсети и настраивает сеть.

Сетевой мост (network bridge) – программно-аппаратный комплекс, объединяющий несколько физических сегментов сети в один логический.

2. Все HTTP(S)-запросы, поступающие от пользователей, через правила ebtables и iptables (утилиты, управляющие внутренним firewall-ом Linux) перенаправляются в прокси-сервер Squid.

3. Все HTTP(S)-запросы Squid посылает по протоколу ICAP REQMOD-запрос в ICAP-сервер regexp_icap_filter. Regexp_icap_filter передает URL запроса и дополнительные данные в content_filter_analyzer.
4. Content_filter_analyzer определяет пользователя и его настройки фильтрации. В зависимости от настроек и категории страницы запрос на данный URL либо разрешается для дальнейшей проверки, либо блокируется.
5. Если запрос блокируется, то пользователю показывается страница блокировки, иначе Squid отправляет запрос дальше к origin серверу.
6. Ответ origin-сервера Squid передает в RESPMOD-запросе в regexp_icap_filter. Сервис regexp_icap_filter собирает весь контент страницы в файл на диске в папке /tmp/ и передает путь до файла в content_filter_analyzer.
7. Сервис content_filter_analyzer считывает текст страницы из файла, проводит синтаксический разбор HTML, выделяет весь текст из страницы. Для каждого слова из текста страницы проводится морфологический анализ принадлежности слова определенным группам слов. Определенным словам и фразам соответствует некоторый вес. Все веса найденных фраз суммируются и считается за вес страницы. При превышении веса страницы определенного веса страница блокируется. Группы слов, вес каждого слова и минимальный вес необходимый для блокировки html страницы настраивается администратором системы.
8. В зависимости от анализа контента страницы пользователю отсылается либо страница блокировки, либо regexp_icap_filter посылает обратно Squid-у оригинальный контент страницы.

2. 2. Описание веб-интерфейса

Управление происходит из веб-интерфейса, который доступен по IP-адресу. Веб-интерфейс является SPA-приложением, сделанный на фреймворке React. Веб-интерфейс доступен как на русском, так и на английском языке. Для того чтобы зайти в веб-интерфейс нужно сначала авторизоваться.

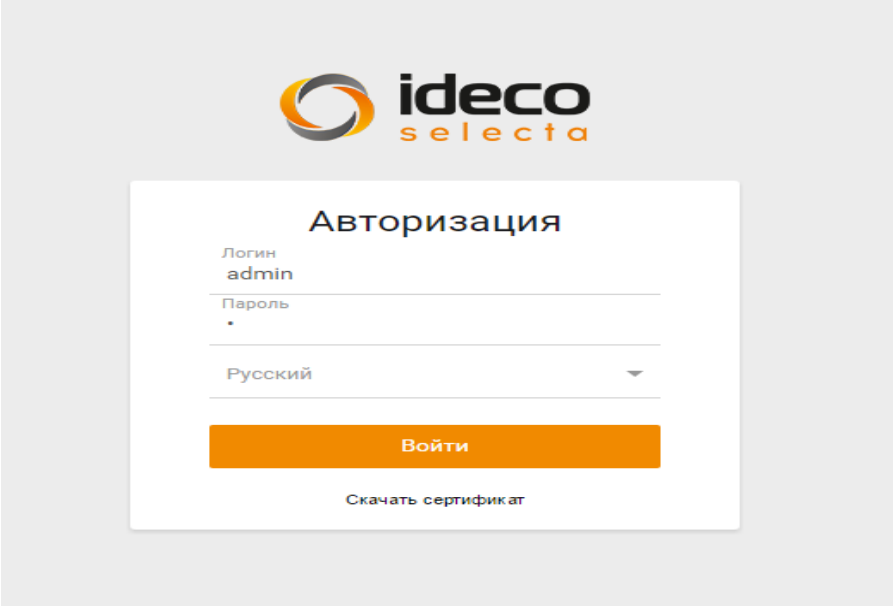


Рис. 2.1. Страница авторизации

На главной страничке расположена информационная панель – dashboard, отражающая основные характеристики состояния сервера: время непрерывной работы, загрузка CPU и памяти системы, текущая пропускная способность системы.

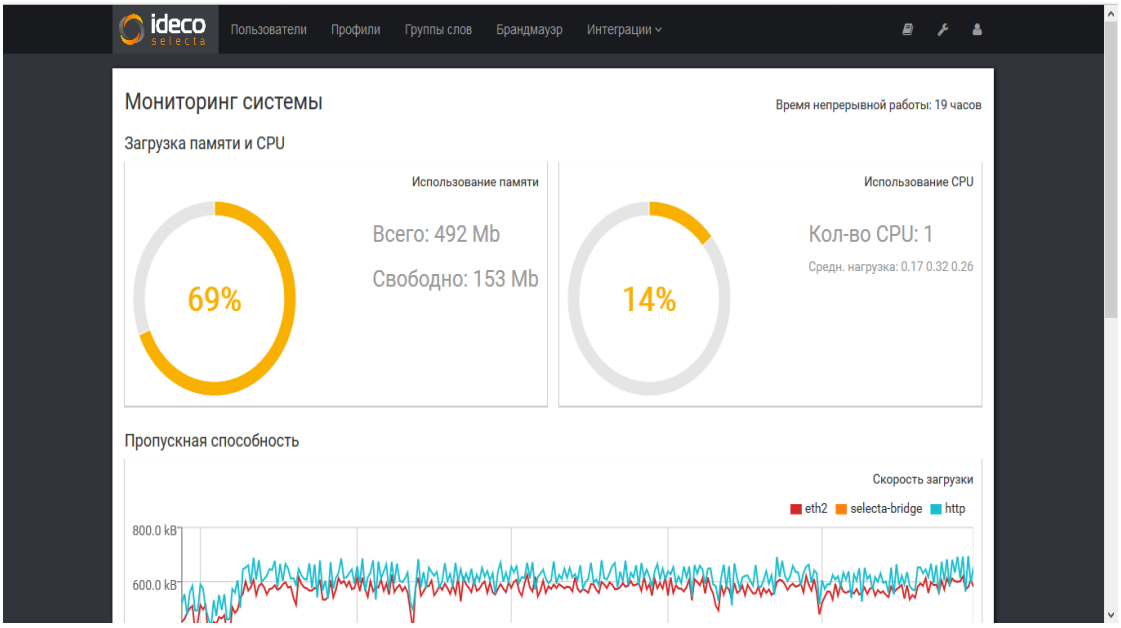


Рис. 2.2. Информационная панель (Dashboard)

Сверху расположено меню с вкладками основными вкладками:

- **Пользователи** – управление пользователями «Selecta», траффик которых будет фильтроваться.
- **Профили** – управление различными настройками фильтрации.
- **Группы слов** – задание различных слов для фильтрации.
- **Брандмауэр** – на этой вкладке администратор может добавить дополнительные правила фильтрации.
- **Интеграции** – различные дополнительные способы интеграции «Selecta» в сеть клиента. А именно:
  - *Active Directory* – добавление «Selecta» в Домен.
  - *ICAP* – добавляет возможность интеграции с другими ICAP серверами. Можно использовать, к примеру, для интеграции с антивирусами (ClamAV, Антивирус Касперского).
  - *SMPP* – добавляет авторизацию пользователей через SMS.

- *WCCP* – интеграция с роутерами через WCCP. Этот способ интеграции в основном нужен для операторов.

2.2.1. Вкладка «Пользователи»

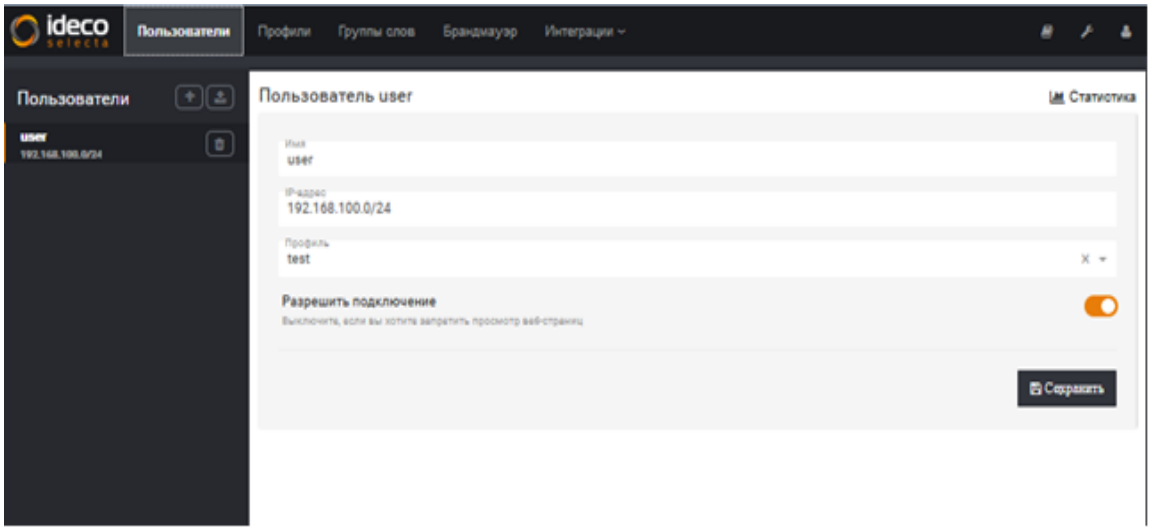


Рис. 2.3. Вкладка «Пользователи»

На вкладке Пользователи задаются клиенты, для которых будет происходить фильтрация трафика. Пользователем может быть, как один IP-адрес, так и целая подсеть в короткой форме (IP-адрес/короткая маска подсети, например, 192.168.1.0/24), для которого также задается имя и профиль, по которому будет фильтроваться запрос пользователя. Если профиль не задан, то в зависимости от переключателя «Разрешить подключение» запросы будут блокироваться, либо пропускаться без фильтрации. Для каждого пользователя можно посмотреть статистику – самые популярные заблокированные и не заблокированные категории, заблокированные и не заблокированные домены. Администратор может загрузить пользователей из файла, содержащий список IP-адресов и подсетей, в таком случае имя пользователя будет совпадать с указанным IP-адресом или подсетью.

2.2.2. Вкладка «Профили»

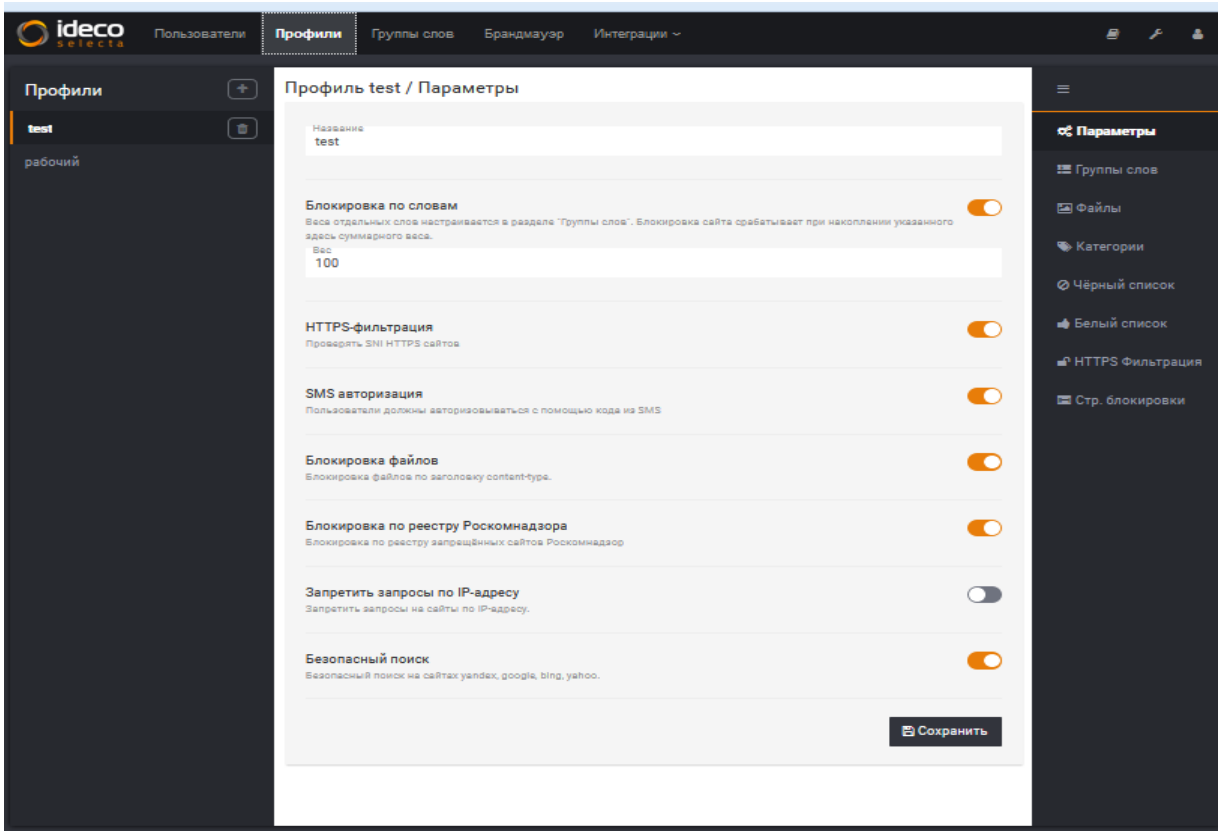


Рис. 2.4. Вкладка «Профили»

На вкладке администратор включает необходимые типы фильтров, определяющих блокировку.

- Подвкладка **Группы слов** позволяет администратору выбрать такие группы слов, по которым будут фильтроваться запросы. Для данного профиля будут блокироваться страницы на которых присутствуют слова из выбранных группы слов.
- Подвкладка **Файлы** - администратор указывает типы файлов, запрещенных для пересылки.
- Подвкладка **Категории** - администратор может выбрать категории сайтов из предложенных, по которым будут блокироваться сайты. Указав в поисковой строке подвкладки страницы сайта, администратор определит категории данного сайта и выберет необходимые для фильтрации.

Подвкладки Черный список, Белый список и HTTPS Фильтрация администратор может указать весь URL-адрес либо шаблон поиска. Строка может содержать символы-джокеры: «?» обозначающее один любой символ и «*» для нуль и более любых символов.

- Подвкладка **Черный список** задаётся шаблон, какие сайты блокировать без дополнительной фильтрации.
- Подвкладка **Белый список** задаётся шаблон, какие сайты пропускать без дополнительной фильтрации.

Если сайт подпадает под шаблон **HTTPS-фильтрации**, то для HTTPS-запроса подменяется оригинальный сертификат на самоподписанный.

- Подвкладка **Страница блокировки** - администратор может создать специализированную страницу блокировки, которая будет показываться вместо исходной html-страницы. Если нажать на ссылку «Вставить описание причины блокировки» или «Вставить найденное содержимое», то на месте курсора добавляется шаблон, где будет текст причины блокировки, либо найденное на странице содержимое.

2.2.3. Вкладка «Группы слов»

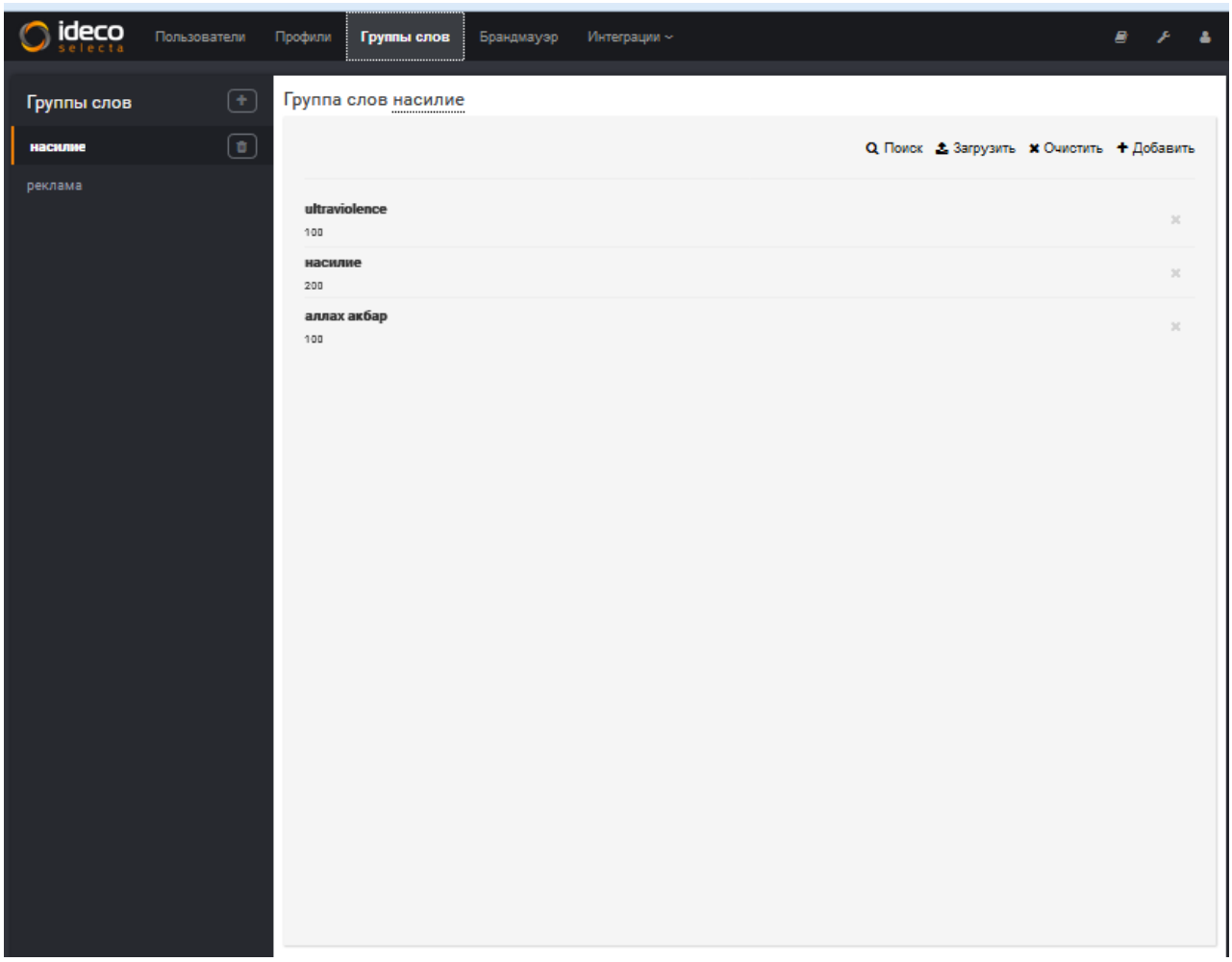


Рис. 2.5. Вкладка «Группы слов»

На вкладке группы слов администратор самостоятельно создает группы слов для блокировки. Наполнение группы возможно, как по одному слову, так и загрузкой слов из файла, включающего несколько слов. Так же необходимо указать вес каждого слова для блокировки. При этом важно учесть, что вес для блокировки каждому загружаемому слову из файла задается автоматически и равен 100.

2.2.4. Вкладка «Брандмауэр»

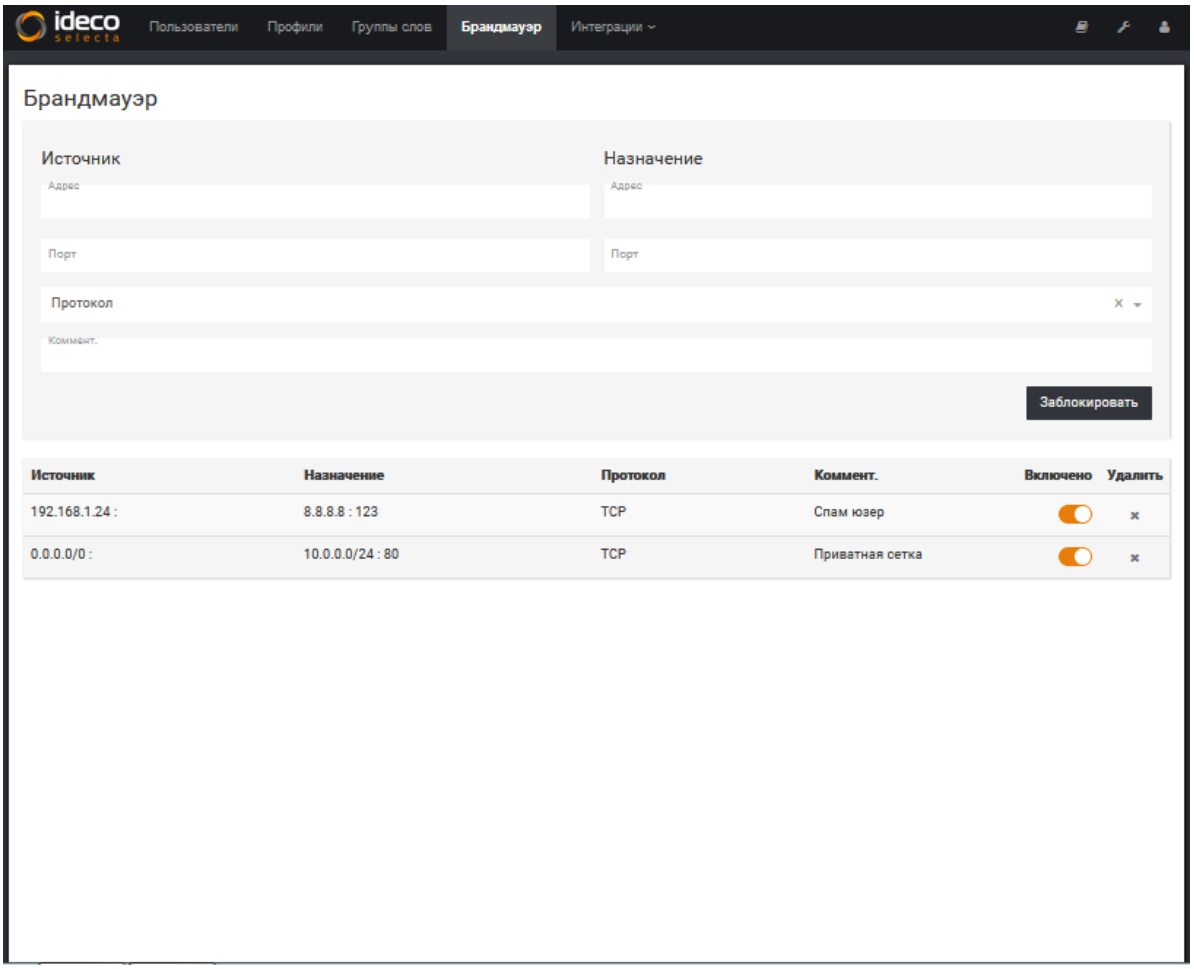


Рис. 2.6. Вкладка «Брандмауэр»

Вкладка Брандмауэр дает администратору минималистичный способ указать правила для файерволла «Selecta». Правило состоит из IP-адреса источника, IP-адреса назначения и одного из протоколов транспортного уровня (TCP, UDP и т.д.). Также можно указать порт источника и назначения и комментарий для облегчения дальнейшей поддержки правил администратором. Список правил отображается на этой же вкладке. Правила применяются в том же порядке как указаны в списке.

2.2.5. Вкладка «Учетные записи»

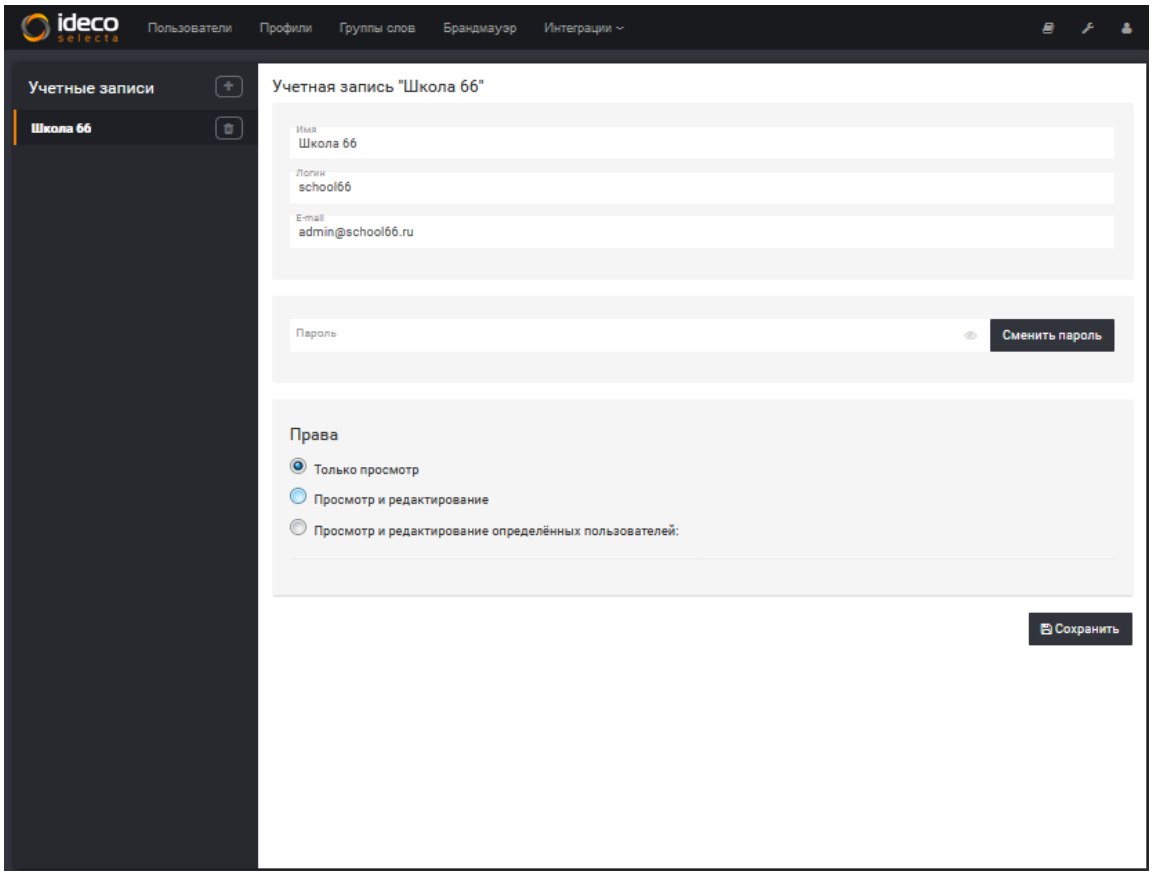


Рис. 2.7. Вкладка «Учетные записи»

Доступ к веб-интерфейсу имеют администраторы с различным уровнем прав. Предусмотрены 3 различных уровней прав.

- *Просмотр и редактирование* – администратор может править все настройки, для всех пользователей, полный доступ.
- *Просмотр и редактирование определенных пользователей* – администратор может изменять настройки только определенных пользователей и просматривать веб-интерфейс.
- *Только просмотр* – администратор может только просматривать веб-интерфейс. Никаких настроек администратор не может изменить.

Первоначально, когда ещё не создано никаких аккаунтов, доступ в веб-интерфейс можно зайти под логином *admin* и любым паролем. Когда создан аккаунт с правами на **просмотр и редактирование**, то зайти под временным аккаунтом не получится.

2.2.6. Вкладка «О продукте»

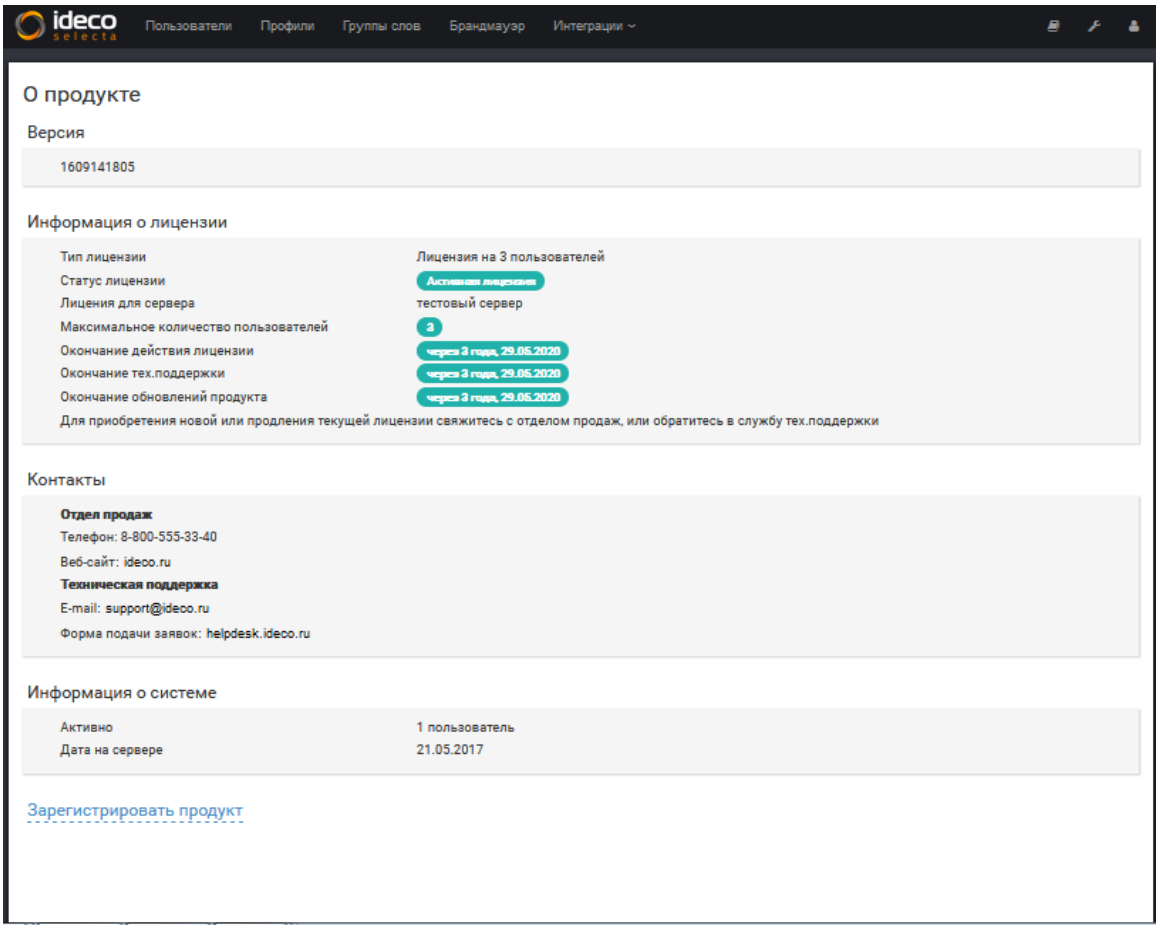


Рис. 2.8. Вкладка «О продукте»

На этой вкладке указана актуальная информация о версии, лицензии, и контактная информация как связаться с компанией Ideco. Также на этой вкладке можно зарегистрировать продукт «Selecta», что необходимо для нормальной работы «Selecta».

2.2.7. Вкладка «Журнал»

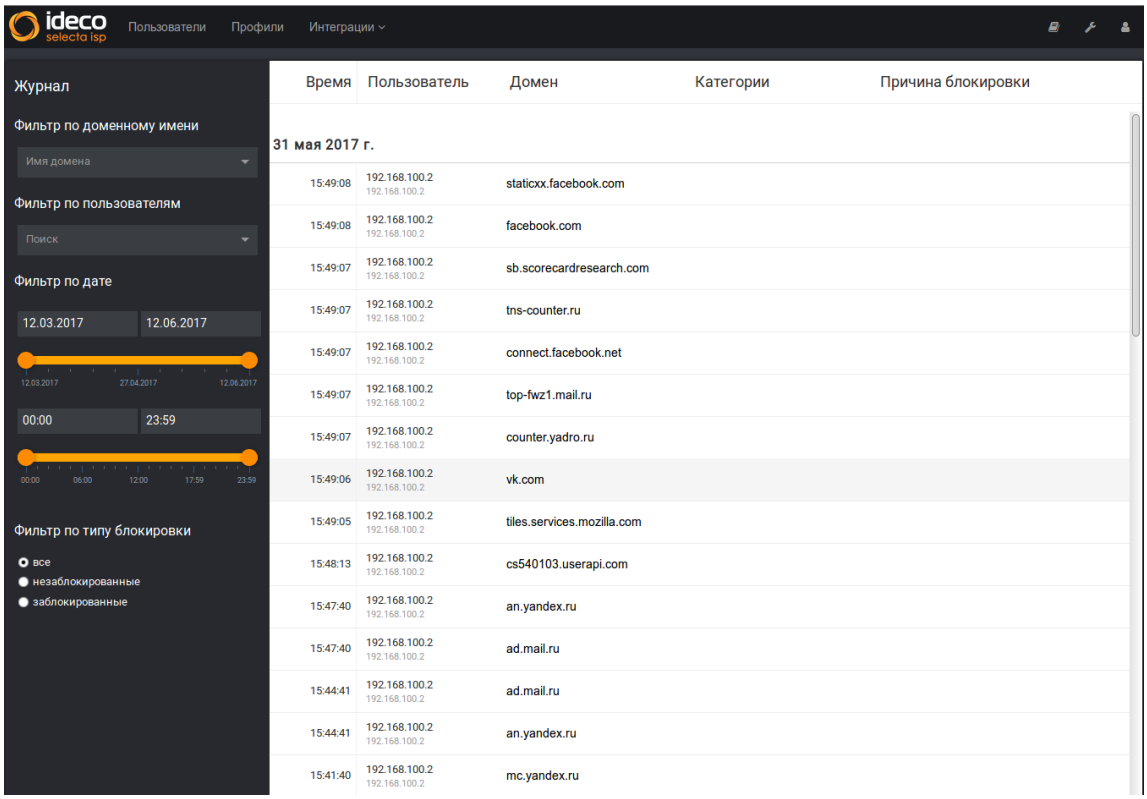


Рис. 2.9. Вкладка «Журнал»

Вкладка **Журнал** нужна для отслеживания работы «Selecta» и просмотра запросов пользователей. Администратор может посмотреть запросы пользователей за определенный промежуток времени для всех или определенных пользователей или посмотреть, кто из пользователей заходил на определенные домены. Кроме того, есть фильтр по типу блокировки:

- показать только все запросы;
- только незаблокированные;
- заблокированные по какому-либо типу.

2.2.8. Вкладка «Настройки»

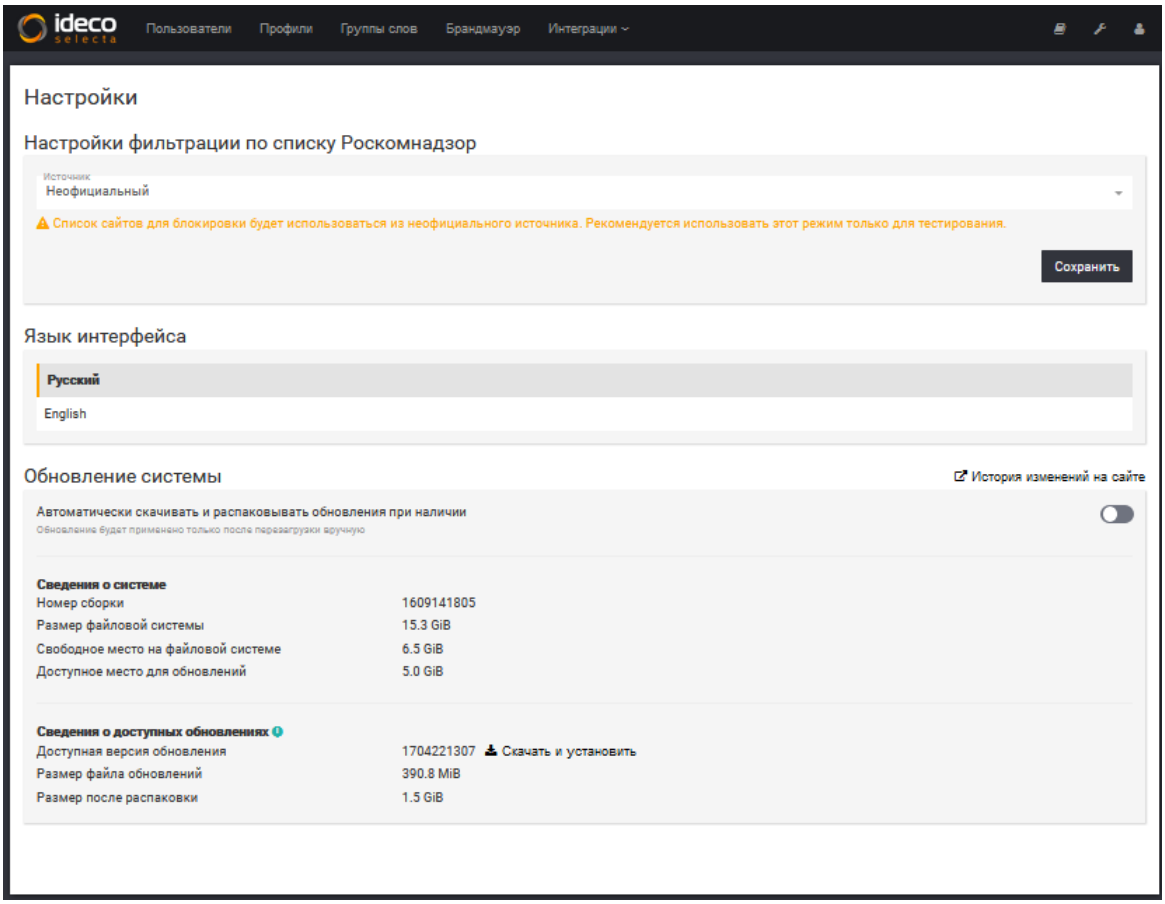


Рис. 2.10. Вкладка «Настройки»

На вкладке **Настройки** расположены основные настройки «Selecta»: язык локализации, выбор источника баз, данных РКН-а и параметры обновления системы. На данный момент «Selecta» поддерживает русский и английский язык.

2. 3. Результаты апробации, техническая документация

Проект «Selecta» был успешно внедрен в сетях образовательных учреждений республики Бурятия и получены положительные отзывы о его работе.

Техническая документация представлена руководством по эксплуатации системы для администраторов (См. файл Руководство.pdf).

Заключение

В соответствии с целью и задачами, сформулированными в квалификационной работе, было проделано следующее:

- рассмотрены локальные нормативные акты и методические материалы для обеспечения информационной безопасности детей при использовании ресурсов сети "Интернет";
- рассмотрены существующие методы фильтрации трафика;
- рассмотрены способы организации контентной системы в образовательном учреждении;
- сформулировано техническое задание;
- разработана и протестирована система контентного анализа «Selecta»;
- данная система была внедрена в учебный процесс республики Бурятия.

Следовательно, можно утверждать, что цели выпускной квалификационной работы были достигнуты, задачи выполнены в полном объеме.

Список литературы

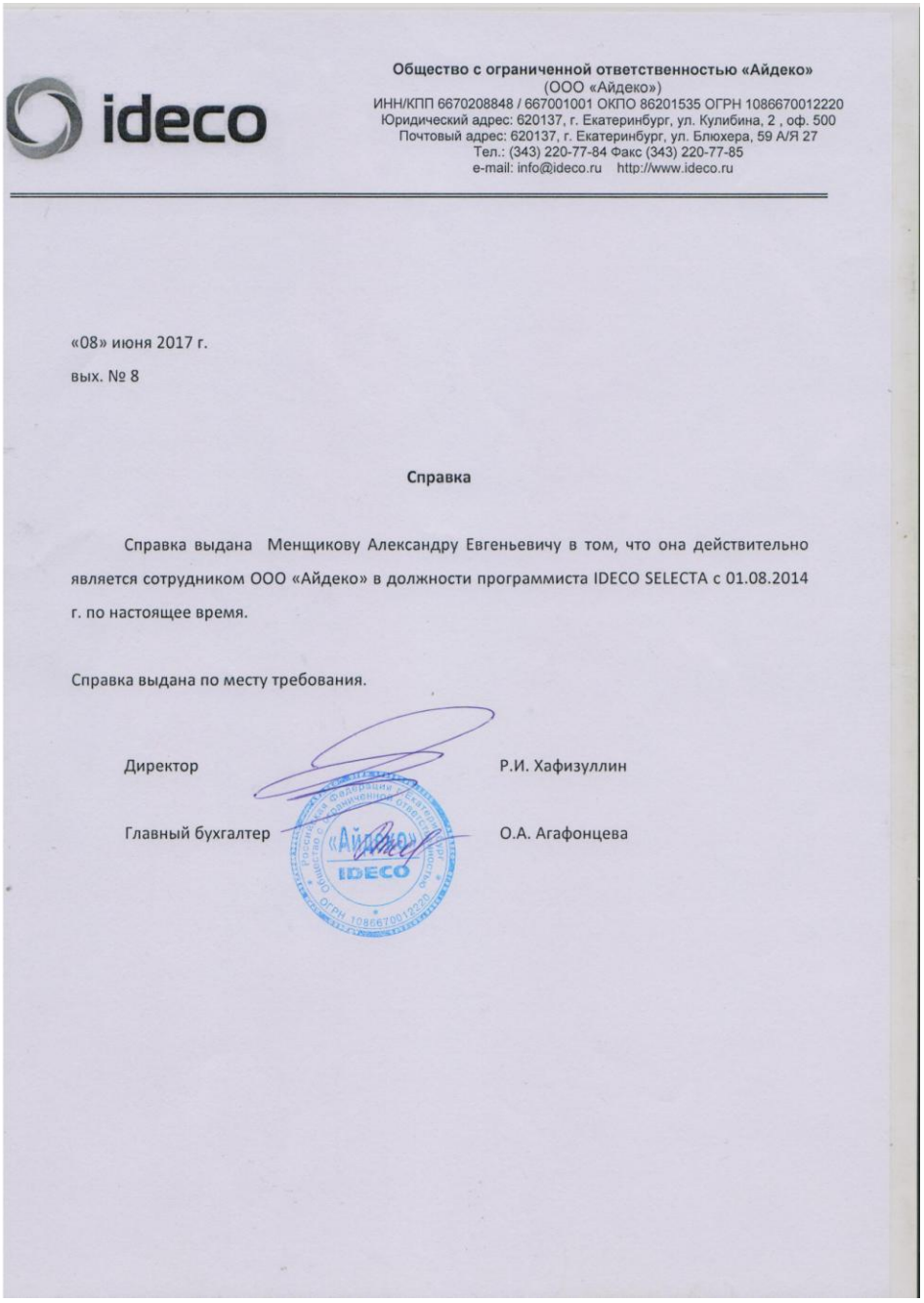
1. Википедия. ARPANET // URL: https://ru.wikipedia.org/wiki/ARPANET (дата обращения: 15.05.2017)
2. Википедия. Всемирная паутина // URL: https://ru.wikipedia.org/wiki/Всемирная_паутина (дата обращения: 15.05.2017)
3. FAQ по Gopher // URL: https://gopherproxy.meulie.net/gopher.viste-family.net/0/gopher-faq/gopher-faq-2009-02-07.txt (дата обращения: 15.05.2017)
4. Википедия. URI // URL: https://ru.wikipedia.org/wiki/URI (дата обращения: 15.05.2017)
5. Википедия. Интернет в КНДР // URL: https://ru.wikipedia.org/wiki/Интернет_в_КНДР (дата обращения: 15.05.2017)
6. Википедия. Кванмён // URL: https://ru.wikipedia.org/wiki/Кванмён_(сеть) (дата обращения: 15.05.2017)
7. Википедия. Great Firewall // URL: https://en.wikipedia.org/wiki/Great_Firewall (дата обращения: 15.05.2017)
8. Википедия. Internet censorship in China // URL: https://en.wikipedia.org/wiki/Internet_censorship_in_China (дата обращения: 15.05.2017)
9. Как ограничивают Интернет в разных странах // URL: http://politmix.ru/content/kak-ogranichivayut-internet-v-raznykh-stranakh (дата обращения: 15.05.2017)

10. Википедия. PIPA // URL: https://ru.wikipedia.org/wiki/PROTECT_IP_Act (дата обращения: 15.05.2017)
11. Российская Газета. Федеральный закон от 27 июля 2006 г. N 149-ФЗ Об информации, информационных технологиях и о защите информации // URL: https://rg.ru/2006/07/29/informacia-dok.html (дата обращения: 22.05.2017)
12. КонсультантПлюс. Федеральный закон "О защите детей от информации, причиняющей вред их здоровью и развитию" от 29.12.2010 N 436-ФЗ (последняя редакция) // URL: http://www.consultant.ru/document/cons_doc_LAW_108808/ (дата обращения: 15.05.2017)
13. КонсультантПлюс. Федеральный закон № 139 «О внесении изменений в федеральный закон «О защите детей от информации, причиняющей вред их здоровью и развитию» от 28.07.12 // URL: http://www.consultant.ru/document/cons_doc_LAW_133282/ (дата обращения: 15.05.2017)
14. КонсультантПлюс. Федеральный закон № 185 от 02.07.2013 // URL: http://www.consultant.ru/document/cons_doc_LAW_148576/ (дата обращения: 15.05.2017)
15. КонсультантПлюс. Федеральный закона № 307 от 14.10.2014 // URL: http://www.consultant.ru/document/cons_doc_LAW_169745/ (дата обращения: 15.05.2017)
16. Википедия. Анонимность в Интернете // URL: https://ru.wikipedia.org/wiki/Анонимность_в_Интернете (дата обращения: 22.05.2017)
17. Киберпсихология: эффект растормаживания в Интернете // URL: http://www.vigonsky.ru/post285954206/ (дата обращения: 22.05.2017)

18. Конституция РФ // URL: http://www.constitution.ru/10003000/10003000-4.htm (дата обращения: 22.05.2017)
19. Методы анонимности в сети // URL: https://habrahabr.ru/post/190396/ (дата обращения: 22.05.2017)
20. Викиучебник. Защита конфиденциальных данных и аннонимность в интернете// URL: https://ru.wikibooks.org/wiki/Защита_конфиденциальных_данных_и_анонимность_в_интернете (дата обращения: 22.05.2017)
21. Туннелирование - взгляд изнутри // URL: http://www.nestor.minsk.by/sr/2003/04/30417.html (дата обращения: 22.05.2017)
22. Википедия. VPN // URL: https://ru.wikipedia.org/wiki/VPN (дата обращения: 22.05.2017)
23. Википедия. Tor // URL: https://ru.wikipedia.org/wiki/Tor (дата обращения: 22.05.2017)
24. Википедия. Луковая Маршрутизация // URL: https://ru.wikipedia.org/wiki/Луковая_маршрутизация (дата обращения: 22.05.2017)
25. Раскрываем секреты сети I2P // URL: https://xakep.ru/2014/09/04/i2p-secrets/ (дата обращения: 22.05.2017)
26. Википедия. I2P // URL: https://ru.wikipedia.org/wiki/I2P / (дата обращения: 22.05.2017)
27. Средства и методы фильтрации контента в интернете // URL: https://sites.google.com/site/metodyblokirovkinezelanojinfor/sredstva-i-metody-filtracii-kontenta-v-internete / (дата обращения: 22.05.2017)
28. «Прозрачный» Squid с фильтрацией HTTPS ресурсов без подмены сертификатов (x86) // URL: https://habrahabr.ru/post/267851/
29. Википедия HTTP // URL: https://ru.wikipedia.org/wiki/HTTP (дата обращения: 22.05.2017)

30. Википедия HTTPS // URL: https://en.wikipedia.org/wiki/HTTPS (дата обращения: 22.05.2017)
31. NetPolice Pro // URL: http://www.netpolice.ru/collection/dlya-ofisa/product/netpolice-pro-litsenziya-na-1-god (дата обращения: 22.05.2017)
32. Сайт NetPolice Pro // URL: http://www.netpolice.ru/ (дата обращения: 22.05.2017)
33. Живой Журнал. Интернет Цензор // URL: http://icensor.livejournal.com/ (дата обращения: 22.05.2017)
34. Интернет Цензор — эффективный родительский контроль // URL: https://vellisa.ru/internet-tsenzor (дата обращения: 22.05.2017)
35. Сайт Traffic Inspector // URL: http://www.smart-soft.ru/products/traffic-inspector/ (дата обращения: 22.05.2017)
36. Habrahabr. Что такое Traffic Inspector и с чем его едят // URL: https://habrahabr.ru/company/smart_soft/blog/225427/ (дата обращения: 22.05.2017)
37. Официальный сайт Squid // URL: http://www.squid-cache.org/ (дата обращения: 22.05.2017)
38. Википедия. SquidGuard // URL: https://en.wikipedia.org/wiki/SquidGuard (дата обращения: 22.05.2017)
39. Cleanfeed // URL: https://www.cybertip.ca/app/en/projects-cleanfeed (дата обращения: 22.05.2017)
40. Википедия. Атака посредника. // URL: https://ru.wikipedia.org/wiki/Атака_посредника SquidGuard (дата обращения: 22.05.2017)
41. Википедия. Server_Name_Indication // URL: https://ru.wikipedia.org/wiki/Server_Name_Indication (дата обращения: 22.05.2017)

Приложение



Общество с ограниченной ответственностью «Айдеко»
(ООО «Айдеко»)
ИНН/КПП 6670208848 / 667001001 ОКПО 86201535 ОГРН 1086670012220
Юридический адрес: 620137, г. Екатеринбург, ул. Кулибина, 2 , оф. 500
Почтовый адрес: 620137, г. Екатеринбург, ул. Блюхера, 59 А/Я 27
Тел.: (343) 220-77-84 Факс (343) 220-77-85
e-mail: info@ideco.ru http://www.ideco.ru

«08» июня 2017 г.

вых. № 8

Справка

Справка выдана Менщикову Александру Евгеньевичу в том, что она действительно является сотрудником ООО «Айдеко» в должности программиста IDECO SELECTA с 01.08.2014 г. по настоящее время.

Справка выдана по месту требования.

Директор Р.И. Хафизуллин

Главный бухгалтер О.А. Агафонцева