Контентная фильтрация в Linux (Ubuntu 10.04)

Автор:
Орлов Дмитрий Владимирович, учитель информатики МОУ Туношенской СОШ.

При написании данной статьи были использованы следующие ресурсы:
Рецепты приготовления кальмара
Настройка Squid для начинающих (squid proxy acl)
Руководство по настройке контент-фильтра Dansguardian
Ubuntu Server. Настраиваем контент-фильтр роутера (DansGuardian)
DansGuardian. Сложности фильтрации русскоязычного контента

Часть 1. Настройка SQUID

Если требуется предоставить совместный доступ к Интернету с возможностью кэширования трафика, то в первую очередь стоит рассмотреть такой вариант, как использование кэширующего прокси-сервера Squid. Это очень гибкое решение, которое применяют как в малых офисах с несколькими пользователями, так и в корпоративных сетях со сложной топологией. Рассмотрим, как установить и настроить прокси-сервер Squid в Ubuntu Linux.

Пошаговая инструкция

1. Установку squid запускаем командой:

```
sudo apt-get install squid3
```

2. Начинаем самый ответственный этап — настройку, т. е. редактирование конфигурационного файла:

```
sudo gedit /etc/squid3/squid.conf
```

3. В открывшемся окне конфигурационного файла находим параметр visible_hostname. Этот параметр устанавливает имя прокси-сервера. Например:

```
visible_hostname mysquid
```

4. Затем, находим параметр ***http_port 3128***, и меняем запись по следующему формату:

```
http_port 192.168.0.1:3128
```

   где прописывается ip-адрес компьютера, на котором собственно и настраиваем squid.

5. В разделе ***ACCESS CONTROLS*** находим строку ***acl all src all*** и настраиваем эту и ближайшие строки так, как показано ниже:

```
    #Default:
    # acl all src all
    #
    #Recommended minimum configuration:
    acl manager proto cache_object
    acl localhost src 127.0.0.1/32
    acl to_localhost dst 127.0.0.0/8 0.0.0.0/32
```

6. Следующий шаг это добавление ip-адресов компьютеров, которые должен обслуживать squid. Например, так (записываются эти строчки сразу после предыдущих):

```
    acl k_100 src 192.168.0.100
    acl k_101 src 192.168.0.101
    acl k_102 src 192.168.0.102
    acl k_103 src 192.168.0.103
    acl k_104 src 192.168.0.104
    acl k_105 src 192.168.0.105
    acl k_106 src 192.168.0.106
    acl k_107 src 192.168.0.107
    acl k_108 src 192.168.0.108
    acl k_109 src 192.168.0.109
```

Ip-адреса, естественно, подставляем свои, реальные. Закоментировав одну и или несколько из этих строк и перезапустив squid мы закроем доступ в Интернет этим компьютерам.

Можно вместо каждого компьютера прописать группу ip-адресов, например, так:

```
acl ucheniki src 192.168.0.101-192.168.0.109
```

7. Пока находимся «в нужном месте» имеет смысл создать правило запрещающее доступ к наиболее популярным ресурсам:

```
acl ucheniki src 192.168.0.101-192.168.0.109
#
#
acl SocSeti dstdomain .vkontakte.ru .vk.com .odnoklassniki.ru
.my.mail.ru .facebook.com
```

Можно дополнить эти адреса самостоятельно.

8. Теперь пришло время воспользоваться теми правилами и соответственно разрешить или запретить доступ в Интернет. Находим ***# TAG: http_access*** и после строки ***#Default***: применяем наши правила:

```
http_access deny ucheniki SocSeti
http_access allow ucheniki
```

deny — запретить, ***allow*** — разрешить. Порядок имеет значение.

9. Запретим всем остальным компьютерам, кроме явно прописанных нами выше, использовать squid для выхода в Интернет, настроив указанный ниже параметр:

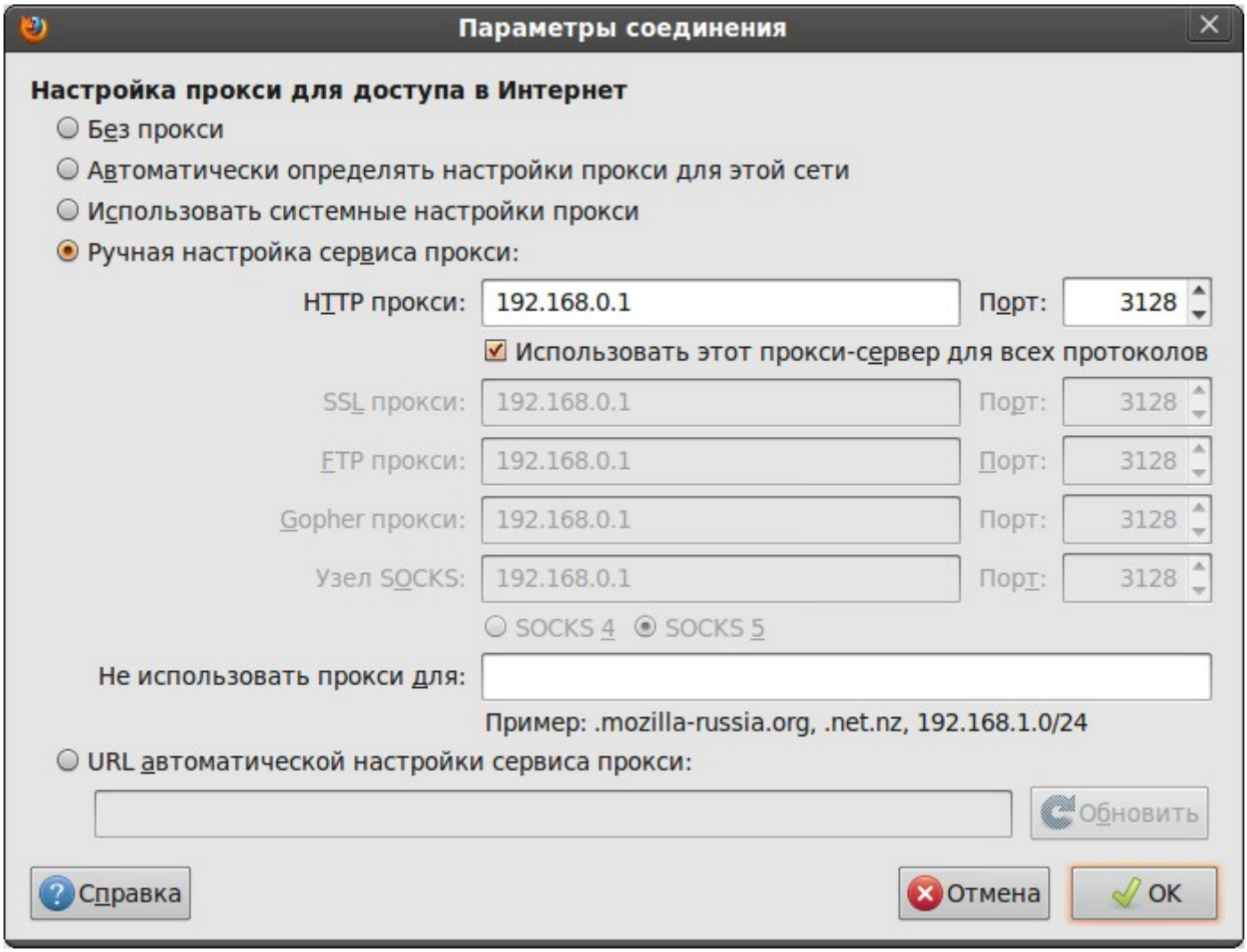
```
http_access deny all
```

10. Дальше сохраняем сделанные настройки, закрываем конфигурационный файл и перезапускаем sqiud:

```
sudo /etc/init.d/squid3 restart
```

11. Проверяем работоспособность. На другом компьютере настраиваем браузер для выхода в Интернет через squid. По умолчанию в Ubuntu установлен Firefox поэтому запускаем его и переходим к настройке Параметров соединения:

```
Правка → Настройки → Дополнительные → Сеть → Настроить
```



12. Применяем, нажав OK и проверяем загрузкой какой-либо страницы Интернет, например: http://tunsh.edu.yar.ru/it_v_shkole.html

Часть 2. Настройка DansGuardian

В определенный момент перед каждым системным администратором встает вопрос фильтрации интернет контента. Особенно актуально этот вопрос стоит в учебных заведениях. Фильтрации по URL и IP в данном случае недостаточно, при таком подходе администратор будет "вечным догоняющим", блокируя нежелательный контент уже после его посещения пользователем. Поэтому наиболее эффективным решением будет установка контент-фильтра.

Одним из лучших решений для платформы Linux, и одним из лучших контент-фильтров вообще, является DansGuardian. Он бесплатен для некоммерческого применения. Для установки контент-фильтра необходим настроенный прокси-сервер Squid.

Приступим к установке настройке контент-фильта Dansguardian.

Пошаговая инструкция

1. Для установки DansGuardian выполним в терминале следующую команду:

```
sudo apt-get install dansguardian
```

2. Теперь откроем конфигурационный файл программы:

```
sudo gedit /etc/dansguardian/dansguardian.conf
```

3. Первое что вы должны сделать, это закомментировать (добавить символ # вначале строки) или совсем удалить строку:

```
UNCONFIGURED - ...
```

4. Следующий параметр **reportinglevel** задает уровень фильтрации, он может принимать значения:

   -1 - Скрытый режим, страницы не блокируются, но ведется лог,
   0 - Выводится "Доступ заблокирован",
   1 - Выводится страница без показа запрещенных фраз,
   2 - Полный отчет
   3 - Выводится HTML шаблон страницы запрета.

По умолчанию уже установлено значение:

```
reportinglevel = 3
```

5. Следующий интересующий нас параметр, это язык HTML шаблона страницы запрета:

```
language = 'russian-koi8-r'
```

6. Теперь раскомментируем строку указывающую путь к файлу лога:

```
loglocation = '/var/log/dansguardian/access.log'
```

7. Укажем сетевой интерфейс и порт на котором DansGuardian будет принимать соединения от клиентов:

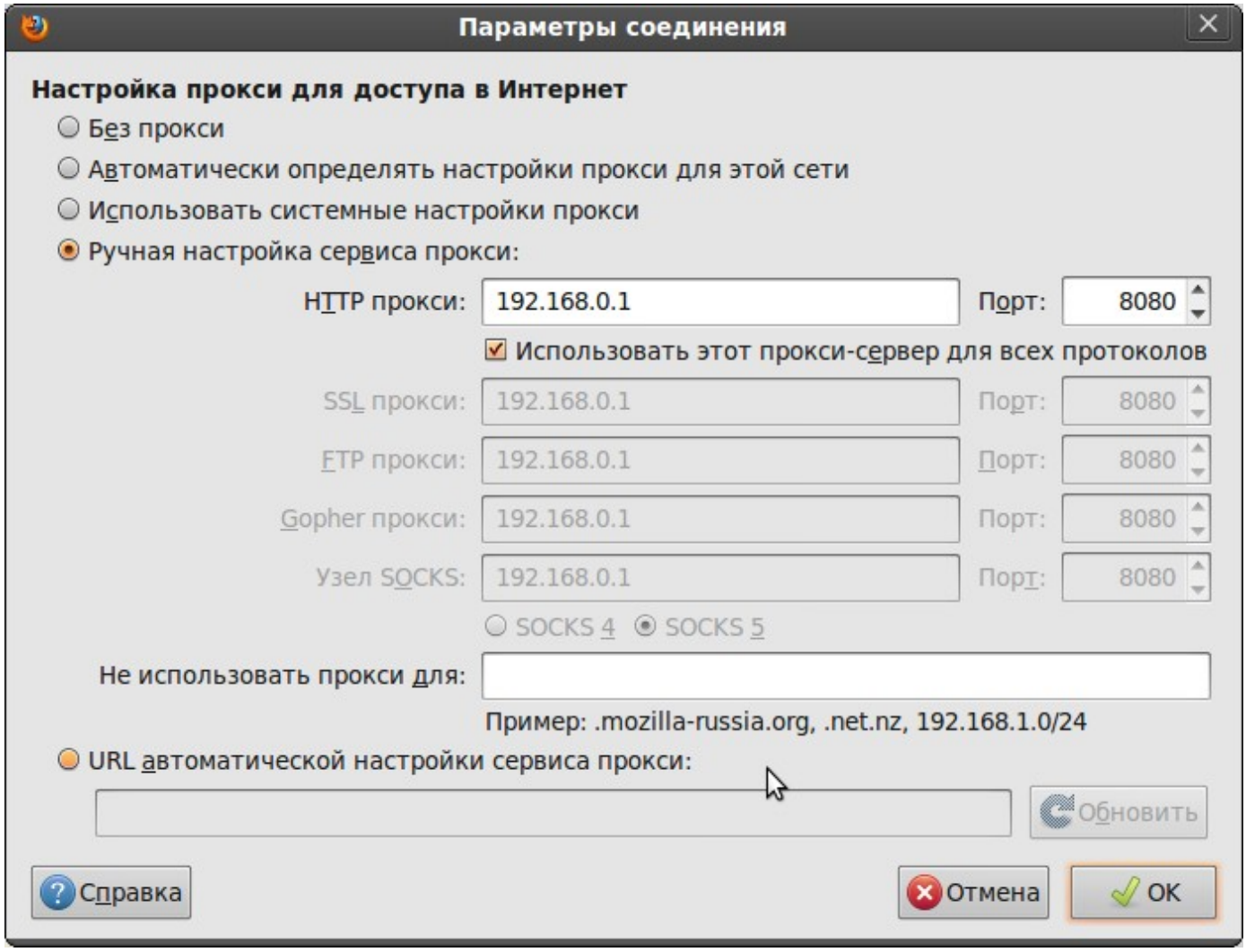
```
filterip = 192.168.0.1
filterport = 8080
```

Остальные параметры менять не нужно. DansGuardian будет работать на порте 8080, в свою очередь Squid должен использовать порт 3128.

8. Сохраняем конфигурационный файл и запускаем DansGuardian:

```
sudo /etc/init.d/dansguardian start
```

9. Для проверки настроим браузер клиентского компьютера на использование прокси-сервера 192.168.0.1:8080



10. Далее проведем более тонкую настройку, которая позволит DansGuardian наиболее эффективно фильтровать русскоязычный контент. Откроем ***/etc/dansguardian/dansguardian.conf*** в котором найдем и установим следующие значения параметров

```
preservecase = 2
hexdecodecontent = on
forcequicksearch = on
```

11. Следующим шагом будет составление списков фраз. Готовить списки нам придется сразу в трех

кодировках (UTF8, Windows-1251, KOI8-R). Перейдем в каталог ***/etc/dansguardian/phraselists*** и создадим там папку для наших списков:

```
cd /etc/dansguardian/phraselists
sudo mkdir rus
```

12. Перейдем в эту папку и создадим заготовки для файлов списка:

```
cd rus
touch rus1251 rusutf8 ruskoi8
```

13. Так как системной консолью Ubuntu является UTF-8 основная работа у нас будет происходить с файлом ***rusutf8***. Добавим в него необходимые фразы и их вес. Формат записи следующий:

```
<фраза>,<40>
```

Остановимся на одном тонком моменте, пробелах перед и после фразы. Их наличие не допускает подстановку символов при проверке. Так например запись < наркотик > будет блокировать слово "наркотик", но не будет "наркотики". А фраза <эро> заблокирует даже такие слова как "аэрофлот", правильнее будет записать < эро>, что будет блокировать все слова с таким началом.

14. Создав список в Юникоде, перекодируем его в остальные кодировки:

```
iconv -f UTF-8 -t WINDOWS-1251 rusutf8 > rus1251
iconv -c -f UTF-8 -t KOI8-R rusutf8 > ruskoi8
```

Попытка перекодировки может закончиться сообщением ***Отказано в доступе***. Тогда необходимо изменить права доступа к файлам ***rus1251*** и ***ruskoi8***.

Это можно сделать командой ***chmod***. Владельца файла обозначает символ 'u', группу владельца - 'g', остальных пользователей - 'o', всех трех типов - 'a'. Установка прав: '+' - добавление права, '-' - удаление права, '=' - явная установка права. Права на чтение - 'r', на запись - 'w', на выполнение - 'x'. Таким образом, чтобы установить всем пользователям права на запись файла ***file***, нужно сделать так:

```
sudo chmod a+w file
```

15. Теперь подключим наши списки. Откроем специальный файл:

```
sudo gedit /etc/dansguardian/lists/weightedphraselist
```

16. В конец файла добавим строки:

```
#Russian
.Include</etc/dansguardian/lists/phraselists/rus/rus1251>
.Include</etc/dansguardian/lists/phraselists/rus/rusutf8>
.Include</etc/dansguardian/lists/phraselists/rus/ruskoi8>
```

17. Сохраняем конфигурационный файл и перезапускаем DansGuardian:

```
sudo /etc/init.d/dansguardian restart
```

Или последовательно:

```
sudo /etc/init.d/dansguardian stop
sudo /etc/init.d/dansguardian start
```