

М. К. Черников

Разработка сенсора, процесса-обработчика спам-статистики и web-страниц на интерфейсе пользователя системы Nadmin

Научный руководитель: сотрудник лаб. «Ботик» А. В. Карлаш

Аннотация. Данная работа освещает проблему борьбы с незапрашиваемой электронной корреспонденцией. В работе рассматривается создание сенсора, процесса-обработчика спам-статистики и web-страниц на интерфейсе пользователя системы Nadmin.

1. Введение

На сегодняшний день не существует действительно надежного метода борьбы с такой головной болью, как незапрашиваемая электронная корреспонденция, или спам [1,2]. Этой проблемой озабочены все: как пользователи, так и поставщики услуг, вплоть до AOL¹ и других крупнейших компаний. Ложное срабатывание — это беда всех антиспам-фильтров. Большие потери несут, конечно, сами пользователи электронной почты, поскольку на сервере теряются важные письма, так и не увидевшие своего получателя. Нынешние антиспам-фильтры не всегда корректно обрабатывают входящую почту, поэтому некоторые письма попадают в корзину для спама. Однако, если проанализировать информацию о письмах, которые были определены как спам, то можно, исходя из этой информации, составить списки «белых» и «черных» адресов и более комфортно управлять настройками своей почты.

¹AOL LLC (бывшая America Online, Inc.) — американская медиа-компания, поставщик онлайн-сервисов и электронных досок объявлений, владелец четвертой по популярности в мире поисковой системы, социальной сети Bebo, популярных интернет-пейджеров ICQ и AIM, а также медиаплеера Winamp.

2. Постановка задачи

Наша цель — это создание механизма обработки и просмотра спам-статистики для дальнейшего анализа, то есть предоставление абоненту СТ «Ботик» возможности просмотра лога срабатывания антиспам-фильтра. Это позволит авторизованному пользователю видеть информацию о незапрашиваемой корреспонденции и отслеживать, насколько правильно сработал антиспам-фильтр. Чтобы это все воплотить в жизнь, необходимо решить следующие задачи:

- (1) создать сенсор, который обрабатывает логфайл и формирует порции статистики (слайсы);
- (2) создать процесс-обработчик, который принимает данные от сенсора и сохраняет их в архив;
- (3) разработать web-страницы на интерфейсе пользователя системы Nadmin, где можно просмотреть спам-статистику за определенный период.

3. Реализация

Разработка ведется в системе Nadmin [3]. Это объясняется богатым выбором инструментов, наличие которых позволяет нам не создавать такие вещи, как, например, web-интерфейс пользователя.

3.1. Общие принципы реализации сенсоров

Для реализации статистической подсистемы в структуру системы Nadmin введено понятие сенсоров. Сенсоры — это обычные процедурные программы на языке perl. Сенсор является обработчиком «сырой»² статистики. Сенсор берет сырую статистику от конкретного источника, обрабатывает ее и отдает основному процессу-обработчику. Обычно естественным источником статистической информации по использованию ресурсов пользователями через различные серверы системы являются логфайлы этих серверов.

²Сырая статистика появляется из различных источников из «внешнего» (по отношению к Nadmin) мира в момент, когда абоненту оказывается та или иная услуга.

3.2. Спам-сенсор

Новый сенсор должен отвечать следующим требованиям:

- (1) уметь разбираться в конкретном виде логфайла и извлекать необходимые данные;
- (2) формировать порции статистики и передавать их процессу-обработчику;
- (3) сохранять информацию в виде закладки, до какой строки был обработан логфайл.

Антиспам-фильтр формирует свой логфайл, а спам-сенсор берет из него данные, которые являются для него сырой статистикой. Записи в логфайле спам-статистики бывают двух видов: однострочные (независимые) и многострочные (зависимые).

В первом случае вся информация, которая необходима для создания слайса, содержится в одной строке. Во втором случае, чтобы сформировать слайс, необходимо обработать несколько строк. В таких случаях запись имеет следующий вид: в первой строке указаны адреса получателя и отправителя, IP-адрес отправителя, и также дается объяснение, почему данное письмо может быть классифицировано как спам. В следующих строках дается подробная информация, которая, по сути, представляет собой заголовок письма. Данный заголовок включает в себя тему письма (subject), информацию о почтовом клиенте и другие данные. Для некоторых строк необходимо выполнять декодирование, поскольку они находятся в формате MIME.

Рассмотрим порцию статистики.

```
Date1: 20070103121110
Date2: 20070103121910
Host: 198.22.33.18
ID: torrent.botik.ru
Msg_FROM: linengelkemet@engelke.de
Msg_TO: someuser@torrent.botik.ru
Subject: reklama v seti
```

Поля Date1 и Date2 указывают период, в течение которого было получено письмо, поле HOST указывает IP-адрес отправителя, поле ID — это подключение абонента, на которое поступило письмо, поля Msg_FROM и Msg_TO — адреса отправителя и получателя соответственно, поле Subject содержит тему письма и значение данного поля может быть пустым.

Вышеуказанная порция статистики описывает следующее: пользователь someuser подключения torrent.botik.ru за указанный период (2007/01-03 12:11:10-2007/01/03 12:19:10) получил письмо от отправителя linengelkemet@engelke.de с IP-адреса 198.22.33.18.

Еще одним важным механизмом постоянной обработки логфайлов являются закладки (marks). Для спам-сенсора закладкой будет являться прочитанная и обработанная строка логфайла, так что при следующем чтении сенсор начнет обработку со следующей строки.

Сенсор должен отслеживать ротацию логфайлов. Под ротацией понимается процесс закрытия, переименования старого логфайла и создание нового, информация которого относится к следующим суткам. Таким образом, сенсор должен уметь при смене логфайлов открыть указатель на новый файл. Большинство сенсоров отслеживают данное событие как смену индексного описателя файла (inode). В случае изменения inode происходит закрытие старого логфайла и открытие нового.

Спам-сенсор запускается процессом-обработчиком как программа-демон, постоянно находящаяся в рабочем состоянии. После получения очередной порции статистики из логфайла и ее обработки сенсор «засыпает» на некоторое время, по прошествии этого времени сенсор просыпается и процесс повторяется. За счет постоянно открытого указателя на логфайл сенсор всегда помнит, до какого места в логфайле он уже обработал статистику. Таким образом, одна строка никогда не обрабатывается дважды. Дальше порции статистики передаются процессу-обработчику.


3.3. Процесс-обработчик спам-сенсора

Процесс-обработчик создан на основе главного процесса системы Nadmin и построен следующим образом. Сначала происходит идентификация порций статистики, то есть процесс-обработчик определяет, к какому сенсору принадлежит статистика. Затем процесс-обработчик принимает слайсы и для каждого из них по полю ID ищет в справочнике организацию, к которой принадлежит эта статистика. Дальше происходит процесс сохранения полученных данных в архив для каждой организации.

Абонент: БОТИК

Пользователь: tux

Права: USER



ЛС (руб.): 0.00 ● Прогноз (дни): 0

ПБ (руб.): 0.00 Цена 1 Мб (руб.): 2.35

ПС (руб.): 0.00

Абоненту
★ ?

Статистика спам почты за март 2009 года

Вернуться к статистике по месяцам

Период	Подключение	IP-номер отправителя	E-mail отправителя	E-mail получателя	Тема письма
Mar 19, 07:09:48	- radoslav.botik.ru	89.218.234.211	linengelkemet@engelke.de	jurij@radoslav.botik.ru	
Mar 19, 07:09:48	- radoslav.botik.ru	89.218.234.211	linengelkemet@engelke.de	jury@radoslav.botik.ru	
Mar 19, 07:09:48	- radoslav.botik.ru	89.218.234.211	linengelkemet@engelke.de	jura@radoslav.botik.ru	
Mar 19, 07:09:48	- radoslav.botik.ru	89.218.234.211	linengelkemet@engelke.de	jurij@radoslav.botik.ru	
Mar 19, 07:09:48	- radoslav.botik.ru	89.218.234.211	linengelkemet@engelke.de	jury@radoslav.botik.ru	

Рис. 1. Web-страница просмотра спам-статистики

3.4. Web-интерфейс

На интерфейсе пользователя системы Nadmin можно получить информацию о спам-статистике. Каждый авторизованный пользователь системы Nadmin имеет возможность просмотреть содержимое своей спам-корзины, а именно: на web-странице создаются ссылки на те месяца за конкретный год, где есть спам-статистика за данный период. Если спам-статистика отсутствует, то выдается соответствующее сообщение.

4. Результаты

Были разработаны сенсор и процесс-обработчик спам-статистики, созданы web-страницы на интерфейсе пользователя системы Nadmin (рис. 1), отвечающие приведенным выше требованиям, и протестированы на локальной машине. Данные модули можно внедрять в систему Nadmin.

Список литературы

- [1] А. Дилевский; И. Сегалович; Д. Тейблом; Принципы и технические методы работы с незапрашиваемой корреспонденцией (Доступно как: <http://www.spamtest.ru/document.html?context=15932&pubid=27&printdoc=1>). ↑1
- [2] Методы борьбы со спамом (Доступно как: http://www.opennet.ru/base/net/spam_greylis.txt.html). ↑1
- [3] Е. Ермилова; П. Жбанов; А. Карлаш; А. Нестеров; Ю. Шевчук; Nadmin — система администрирования для региональных сетей., 2004. ↑3

М. К. Chernikov. *Sensor, a SPAM statistics process-handler and web-pages in the Nadmin system User Interface development* // Proceedings of Junior research and development conference of Ailamazyan Pereslavl university. — Pereslavl, 2009. — p. 218–223. (*in Russian*).

ABSTRACT. This paper sheds light on the subject of SPAM control. In this work development of a sensor, a SPAM statistics process-handler and web-pages in the Nadmin system User Interface are considered.