УДК 004.3

А. И. Власов, канд. техн. наук, **И. Г. Цыганов**,
МГТУ им. Н. Э. Баумана

Адаптивная фильтрация информационных потоков в корпоративных системах на основе механизма голосования пользователей

Представлен подход к решению задачи адаптивной фильтрации информационных потоков в корпоративных системах на основе механизма голосования пользователей. В задаче выделены два этапа: этап первичной и вторичной фильтрации. Представлена общая структурная схема подхода. Описаны методы фильтрации, применяющиеся на каждом этапе. Приводятся основные математические соотношения, позволяющие автоматически настраивать фильтры в режиме интерактивного голосования пользователей по отдельным сообщениям. Вводится система показателей, позволяющих учесть квалификацию пользователей при голосовании.

Введение

В последнее время отмечается резкое возрастание интереса к защите корпоративных информационных систем (КИС) от нежелательной информации, поступающей из внешних источников. Непродуктивные затраты времени пользователей и повышение требований к техническим средствам КИС, обусловленные потребностью обработки нежелательной информации, ведут к значительным убыткам предприятия. Поэтому сейчас в КИС применяют различные средства фильтрации информационных потоков (ФИП), которые позволяют выявить поступление извне нежелаемой информации и осуществить ее блокировку.

К нежелательной информации относится, например, незапрашиваемая рассылка (НР), передаваемая в глобальной сети Интернет по протоколам электронной почты (ЭП) SMTP, POP3, IMAP4, НР (называемая также англоязычным термином "спам") — это массовая рассылка сообщений (как правило, рекламного характера), инициируемая отправителем без учета потребностей и вопреки желанию адресатов [1, 2]. Число циркулирующих в сети Интернет сообщений НР (СНР) неуклонно растет и уже сейчас составляет более половины от общего числа сообщений ЭП [3, 4]. Поэтому проблема ФИП НР в настоящее время стоит достаточно остро. В данной работе мы будем рассматривать задачу ФИП именно этого типа.

Основные сложности при разработке средств и методов ФИП НР связаны с тем, что распространители НР в целях обхода фильтров постоянно изменяют передаваемые СНР. Это заставляет разра-

батывать гибкие методы фильтрации, допускающие регулярную перенастройку в ответ на ошибки ФИП. Несмотря на имеющиеся отличия, контур настройки в большинстве современных систем фильтрации (*Bogofilter* [5], *SpamAssasin* [6], *Vipul Razor* [7], *SpamTest/Kaspersky AntiSpam* [8]) функционирует сходным образом согласно схеме с обратной связью (рис. 1, см. третью сторону обложки).

В приведенной схеме настройка алгоритмов ФИП осуществляется вручную специально подготовленным персоналом — администраторами. Пользователи, которые первыми выявляют ошибки ФИП, могут только сообщать о них администратору.

Несмотря на распространенность и популярность такого подхода, он имеет важный недостаток, связанный с существенными задержками в контуре настройки фильтров. Наибольшие задержки вызывает звено администратора, время реакции которого может составлять от нескольких часов до нескольких дней и недель (зависит от его загруженности). Такие задержки открывают существенную брешь в корпоративной системе безопасности и в настоящее время во многих случаях являются неприемлемыми.

В связи с этим сейчас наиболее важной задачей является разработка средств фильтрации нового поколения, способных оперативно перенастраиваться в автоматическом режиме в ответ уже на первый сигнал от пользователей об ошибках ФИП. Этот подход позволяет если не исключить полностью, то значительно сократить участие ручного труда администратора в настройке ФИП и значительно повысить эффективность работы систем фильтрации. В данной статье рассматривается комбинированный подход, в котором настройка контуров ФИП может осуществляться как полностью автоматическими методами на основании результатов голосования пользователей по отдельным сообщениям, так и с участием администратора (в ручном режиме).

Общее решение задачи фильтрации

В данной работе задача фильтрации поступающего в КИС сообщения ЭП решается в два этапа: *первичная фильтрация* (ПФ) и *вторичная фильтрация* (ВФ). ПФ реализует общесистемные методы фильтрации, применяемые ко всем поступающим в КИС сообщениям, независимо от их конечного адресата. ВФ настраивается индивидуально отдельными пользователями (рис. 2, см. третью сторону обложки). На каждом этапе проводится:

- выявление в поступившем сообщении характерных для HP признаков;
- принятие на их основе решения о том, является ли данное сообщение CHP;
- действия, обслуживающие логику блокировки сообщения, если имеется уверенность в том, что данное сообщение является CHP.

Сообщение *M*, поступающее из внешних сетей, вначале проходит ПФ. Методы, применяющиеся в ПФ, не зависят от адресата сообщения и реализуют фильтрацию по общим признакам, характерным для HP. Общие признаки формируются на основании коллективного решения отдельных пользователей (по результатам ВФ) и решений администратора системы. В результате сообщение либо отвергается (*S*), либо передается в один или несколько блоков ВФ (*P*), число которых зависит от числа адресатов данного сообщения.

Каждый блок ВФ функционирует от имени одного конкретного пользователя и реализует его индивидуальные настройки. ВФ осуществляется в соответствии с результатами ПФ; дополнительными условиями фильтрации, определяемыми пользователем; результатами голосования других пользователей. ВФ для каждого сообщения позволяет сформировать результат в виде одного из четырех значений:

- *"SA" (Spam Auto)* — сообщение автоматически отнесено к HP;
- *"SM" (Spam Manual)* — пользователь указал вручную, что данное сообщение — HP;
- *"HA" (Ham Auto)* — сообщение автоматически отнесено к обычным сообщениям;
- *"HM" (Ham Manual)* — пользователь вручную указал, что данное сообщение обычное.

Эти значения рассматриваются как результаты голосования и поступают в схемы управления. Схемы управления определяют политику настройки ПФ в соответствии с результатами ВФ и установками администратора.

Методы анализа сообщений

Сообщение рассматривается как сложный информационный объект, состоящий из (рис. 3, см. третью сторону обложки): технического конверта и тела сообщения. Тело сообщения, в свою очередь, делится на заголовки и содержание.

Технический конверт содержит информацию, передающуюся при установлении *SMTP*-соединения (*IP* — адрес удаленной системы, *E-mail* — адрес отправителя и получателя). Заголовки сообщения содержат стандартные форматированные поля

(например, *From*:, *To*: и т. д.). Содержание определяется текстовой частью сообщения.

В процессе ФИП выполняется анализ различных элементов сообщения в целях определения характерных для СНР признаков в каждом из них.

Анализ технического конверта и заголовков осуществляется с помощью методов формального анализа с использованием системы логических условий (*если*...) и соответствующих им действий (... *то* ...).

Для анализа содержания применяются два типа методов: сигнатурный анализ и контекстный анализ.

В основе сигнатурного анализа лежит ведение базы данных (БД) сигнатур всех выявленных СНР. Каждое входящее сообщение проверяется на наличие его в такой БД, что позволяет автоматически отклонять повторный прием ранее выявленных СНР.

Контекстный анализ позволяет определить, является ли данное сообщение СНР на основании вхождения в его текстовую часть характерных слов и словосочетаний. В данной работе задача контекстного анализа решается с помощью многослойной нейросети с переменной структурой [9], синтез которой осуществляется на основании обучающей выборки, в которую заносятся как сообщения СНР, так и обычные пользовательские сообщения. Нейросеть осуществляет взвешенную оценку того, к какому типу относится данное сообщение (СНР или обычное пользовательское).

Необходимо различать методы *жесткой* и *мягкой* фильтрации. Первые позволяют однозначно определить тип сообщения. К таким методам относятся сигнатурный анализ и частично формальный анализ. Настройка жестких методов в ПФ осуществляется только в ручном режиме. Методы мягкой фильтрации позволяют выявить тип сообщения только с некоторой долей уверенности. Уничтожение сообщений на основании методов мягкой фильтрации в блоках ПФ не проводится, однако такие действия могут осуществляться в блоках ВФ в соответствии с индивидуальными настройками пользователя.

Первичная фильтрация информационных потоков

Каждое поступающее на вход сообщение проходит последовательно анализ в трех блоках: сигнатурного анализа, формального анализа и контекстного анализа. Для оценки результатов анализа в каждом из блоков используется целочисленная величина, которую будем называть *уровнем уверенности* (УУ) в том, что данное сообщение является СНР. УУ ассоциируется с каждым сообщением и вычисляется в процессе анализа. УУ будем обозначать G. Величина G оценивается в процентах в диапазоне 0—100. Два крайних значения (0 и 100) используются для индикации полной уверенности в том, что: 0 — сообщение пользовательское; 100 — СНР. Промежуточные величины используются в тех случаях, когда полная уверенность отсутствует. По результатам анализа в каждом из блоков сообщение блокируется только в том случае, когда $G = 100$. Если $G = 0$, то его обработка в блоках ПФ прекращается, и оно передается в блоки ВФ.

Сигнатурный анализ. Структурная схема сигнатурного анализа представлена на рис. 4.

Центральным элементом этой схемы является БД, в которой сохраняются сигнатуры выявленных СНР. Сигнатура является 120-битным числом, рассчитанным с помощью хэш-функции (Nilsimsa), которое представляет содержание сообщения с точностью до незначительных изменений буквенного состава. Каждое поступающее в КИС сообщение кодируется с помощью той же хэш-функции, и в том случае, если полученная сигнатура имеется в БД, то осуществляется блокировка этого сообщения ($G = 100$). Используемая хэш-функция позволяет идентифицировать сообщения как СНР даже в том случае, если в его содержание были намеренно внесены изменения (орфографические ошибки, пропуски букв и т. д.).

Формальный анализ. Если сообщение не заблокировано на этапе сигнатурного анализа, то оно подвергается формальному анализу. Перед началом формального анализа устанавливается текущая величина УУ в $G = 1$ (с большой долей уверенности — пользовательское сообщение). Обработка сообщений в блоке формального анализа осуществляется с помощью последовательного применения к техническому конверту и заголовкам сообщения правил фильтрации. Каждое правило формализуется в виде кортежа $\langle C, A \rangle$. Здесь C — множество условий, A — действия, выполняе-

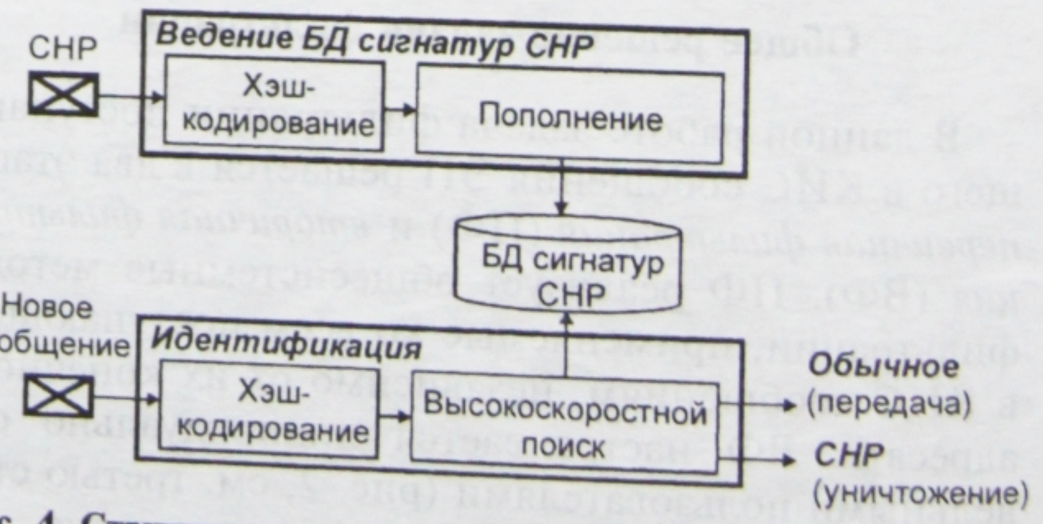

Рис. 4. Структурная схема сигнатурного анализа

мые в случае, если и только если все перечисленные в *C* условия выполнены.

Действия *A* бывают трех типов:

- действия с переменными (определяются новые переменные, изменяются уже имеющиеся);
- действия по изменению УУ *G* текущего сообщения;
- действия по управлению потоком выполнения правил фильтрации (останов, переход и т. д.).

Условия, формирующие компонент *C* каждого правила, представляют собой выражения, результат которых рассматривается как булево значение (*ИСТИННО* или *ЛОЖНО*). Условие может быть представлено как утверждение или отрицание высказываний следующих типов:

- текущий УУ *G* данного сообщения больше (равен, меньше) заданного значения;
- введенная ранее числовая переменная больше (равна или меньше) заданного значения;
- введенная ранее строковая переменная соответствует шаблону;
- IP-адрес SMTP-отправителя находится в заданной БД;
- IP-адрес SMTP-отправителя отсутствует в DNS;
- Email-адрес отправителя находится в заданной БД;
- среди заголовков сообщения имеется заданный заголовок (например, *From*, *To* и т. д.);
- определенный заголовок сообщения соответствует заданному шаблону;
- размер текстовой части сообщения превышает заданное число байтов.

Базы данных, которые могут входить в состав условий, бывают следующих типов:

- локальный черный список IP-адресов (формируется администратором самостоятельно на основании анализа результатов голосования и заявок пользователей);
- локальный белый список IP-адресов (формируется администратором системы на основании анализа заявок пользователей);
- внешние БД RBL (Runtime Blackhole List);
- локальный черный список Email-адресов (формируется администратором самостоятельно на основании анализа результатов голосования и заявок пользователей);
- локальный белый список Email-адресов (формируется администратором системы на основании анализа заявок пользователей).

Правила составляются администратором в ручном режиме и сохраняются в профиле ФА (ПФА).

Общий алгоритм исполнения правил фильтрации приведен на рис. 5. Хотя правила в ПФА могут

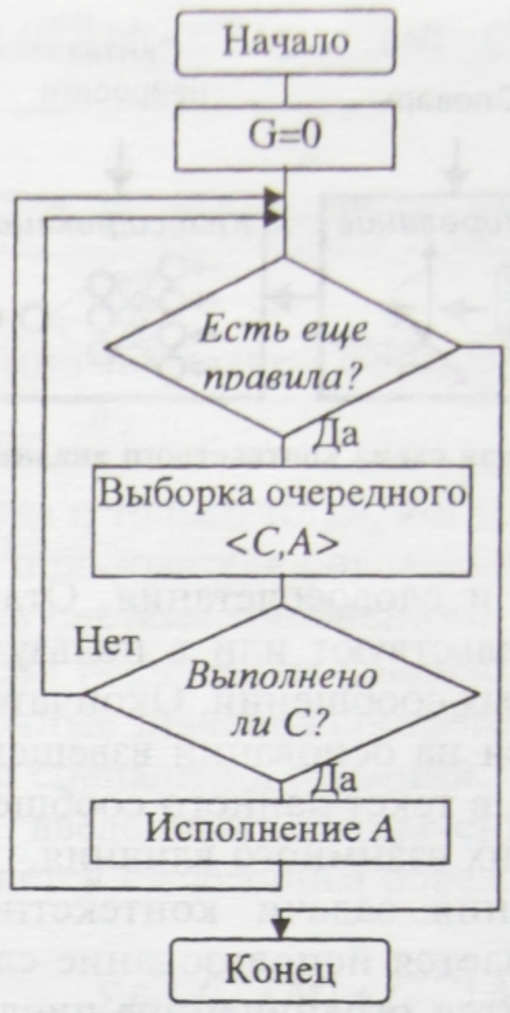

Рис. 5. Обобщенный алгоритм формального анализа

быть приведены в любой последовательности, наиболее предпочтительным является следующий порядок действий:

1. Проверка IP- и Email-адресов:

1.1. Проверка по БД черных списков IP-адресов.

1.2. Проверка по БД белого списка IP-адресов.

1.3. Проверка по БД черного списка Email-адресов отправителей.

1.4. Проверка по БД списка Email-адресов отправителей.

2. Если УУ *G* в результате проверок равен 0 или 100 — переход к п. 6.

3. Проверка заголовков сообщения по шаблонам, характерным для СНР.

4. Проверка IP-адреса отправителя по БД RBL.

5. Проверка того, имеется ли IP-адрес отправителя в БД DNS.

6. В том случае, если текущий уровень $G = 0$, то сообщение передается в блоки ВФ.

7. Если текущий уровень *G* равен 100, то дальнейшая обработка сообщения прекращается и сообщение уничтожается.

8. В противном случае сообщение передается в блок контекстного анализа.

Контекстный анализ. В этом виде анализа решение о том, является ли данное сообщение СНР или это обычное пользовательское сообщение, выносится на основании анализа совокупности признаков, выделяемых из текстовой части сообщения. В качестве признаков в данной работе рассматри-

Рис. 6. Структурная схема контекстного анализа

ваются слова и словосочетания. Отдельные признаки свидетельствуют или в пользу СНР, или в пользу обычных сообщений. Окончательное решение выносится на основании взвешенной оценки совокупности в текст данного сообщения признаков с учетом их взаимного влияния.

Для решения задачи контекстного анализа предусматривается использование словаря, в который заносится ограниченное число (n) наиболее значимых для фильтрации НР признаков (слов и словосочетаний). Все единицы словаря нумеруются, что позволяет закодировать текстовую часть поступающего в контекстный анализ сообщения в виде вектора:

$$X = (x_1, x_2, ..., x_n),$$

где i-я компонента соответствует i-й единице словаря: $x_i = 0$, если i-я единица словаря отсутствует в данном сообщении; $x_i = 1$ — если имеется. Помимо бинарного варианта кодирования вектора сообщения X могут применяться и другие, более сложные модели.

Полученный при кодировании вектор поступает на вход нейросетевого классификатора (рис. 6). Нейросетевой классификатор осуществляет оценку величины G для данного сообщения на основании принадлежности входного вектора к той или иной области n-мерного векторного пространства.

Синтез нейросети и формирование словаря осуществляется в схемах управления (см. рис. 2 на третьей стороне обложки) в автоматическом режиме на основании обучающей выборки.

Вторичная фильтрация информационных потоков

В том случае, если не была проведена блокировка сообщения в блоках ПФ, оно поступает в блоки ВФ. В отличие от ПФ, в ВФ не осуществляется явное выделение отдельных блоков по характеру проводимого анализа. Вместо этого анализ выполняется унифицированным методом на основе правил фильтрации вида $\langle C, A \rangle$.

Действия A здесь следующие:

- признать сообщение СНР и уничтожить его;
- признать сообщение СНР и поместить в указанную папку;
- признать сообщение обычным пользовательским и поместить в указанную папку.

Здесь отсутствуют команды управления потоком выполнения правил. Поэтому проверяются все правила в порядке их указания в профиле фильтрации до тех пор, пока условие хотя бы одного не будет выполнено. Как только такое правило обнаружено, выполняется одно из указанных действий и процесс проверки условий прекращается. Среди условий, которые могут фигурировать в множестве C каждого правила фильтрации, следует выделять *четыре группы*:

- стандартные условия формального анализа;
- условия по результатам ПФ;
- условия совместной фильтрации;
- условия контекстной фильтрации.

Первая группа условий является сокращенным вариантом множества условий, рассмотренных для формального анализа в ПФ. Отличие заключается в том, что групповые условия с использованием БД заменены частными условиями с прямым указанием конкретных значений:

- IP-адрес SMTP-отправителя совпадает с заданным;
- Email-адрес отправителя совпадает с заданным;
- определенный заголовок сообщения соответствует заданному шаблону.

Вторая группа условий фильтрации определяется правилами двух типов:

- УУ, полученный с помощью блока формального анализа ПФ больше (меньше) заданного значения;
- УУ, полученный с помощью блока контекстного анализа ПФ больше (меньше) заданного значения.

Третья группа условий определяется на основании результатов голосования отдельных пользователей по данному сообщению.

Для этого используются группы пользователей, на множестве которых определяются различные операции. Группы могут самостоятельно формировать пользователи на основании личных предпочтений. Для каждого сообщения M_j определяется стандартная группа пользователей, в которую включаются пользователи, получившие это сообщение, которую обозначим $U[M_j]$. Группы могут формироваться из уже существующих на основании простейших теоретико-множественных операций (пересечение, дополнение, объединение и т. д.).

Решение i-го пользователя, получившего j-е сообщение M_j, будем обозначать $S_i[M_j]$ и рассматривать как случайную величину, которая может принимать одно из четырех значений $\{SA, SM, HA, HM\}$. Если из контекста понятно, о каком сообщении идет речь, то решение i-го пользователя будем записывать просто S_i.

Пусть задана группа U, состоящая из h пользователей, решение каждого из которых по данному сообщению известно: $S_i = s_i$, $i = 1, ..., h$. Тогда, если S_0 — случайная величина, соответствующая решению пользователя данного блока, то можно ввести в рассмотрение апостериорную вероятность следующего вида:

$$p(S_0 = s_0 | S_i = s_i, \quad i = 1, ..., n) = \frac{p(S_i = s_i, i = 0, ..., h)}{p(S_i = s_i, i = 1, ..., h)}.$$

Величина $p(S_0 = s_0 | S_i = s_i, i = 1, ..., n)$ показывает вероятность того, что данное сообщение имеет статус s_0 в данном блоке ВФ при условии, что известен его статус, полученный в других блоках ВФ, владельцы которых принадлежат рассматриваемой группе U. Если апостериорную вероятность умножить на 100 и округлить, то данное значение может рассматриваться наравне с УУ. Для его обозначения в работе используется термин *апостериорный уровень группы* (АУГ). В том случае, если группа состоит из единственного пользователя, то рассматривается АУП — *апостериорный уровень пользователя*.

Рассматриваются также оценка *средневзвешенного решения* (СВР) по пользователям, получившим данное сообщение. Для этого вводится четыре индикаторные функции:

$$f_i^S[M_j] = \begin{cases} 1, & \text{если } s_i[M_j] = S; \\ 0, & \text{в противном случае;} \end{cases}$$

$$S \in \{SA, SM, HA, HM\},$$

где i — индекс пользователя; j — индекс сообщения.

Введем также величину квалификации i-го пользователя относительно данного. Эта величина c_i^0 показывает, как часто решение i-го пользователя совпадает с решением данного. Обозначим U_i^0 множество сообщений, которые одновременно получили и i-й пользователь, и данный. Тогда

$$c_i^0 = \frac{\sum_{M \in U_i^0} \left| (f_0^{SM}[M] + f_0^{SA}[M])(f_i^{SM}[M] + f_i^{SA}[M]) \right| + \left| (f_0^{HM}[M] + f_0^{HA}[M])(f_i^{HM}[M] + f_i^{HA}[M]) \right|}{|U_i^0|}.$$

На основании последней формулы можно сделать вывод, что c_i^0:

- равна 1 тогда и только тогда, когда решение i-го и данного пользователя совпадают по всем полученным в системе сообщениям;
- равна 0, если решение не совпало ни разу;
- промежуточные значения соответствуют случаю частичного совпадения решений.

С учетом введенных обозначений величина СВР определяется следующим образом:

$$CBP[M] = \frac{\sum_i c_i^0 \left| f_i^{SA}[M] + f_i^{SM}[M] \right|}{|U[M]|} 100,$$

где M — рассматриваемое сообщение.

В качестве условий в правилах третьей группы могут выступать следующие:

- АУГ определенной группы пользователей больше (меньше) заданного значения;
- АУГ определенного пользователя больше (меньше) заданного значения;
- наиболее вероятное значение для статуса сообщения в данном блоке ВФ при условии, что известен статус данного сообщения в блоках ВФ заданной группы, равен $SA(SM, HA, HM)$;
- наиболее вероятное значение для статуса сообщения в данном блоке ВФ при условии, что известен статус данного сообщения в блоках ВФ заданного пользователя, равен $SA(SM, HA, HM)$;
- величина средневзвешенного решения по пользователям определенной группы больше (меньше) заданного значения.

Четвертая группа условий позволяет осуществить простейшие функции анализа содержания сообщения:

- в произвольной части текста содержится заданная строка;
- в конце текста имеется заданная строка;
- в начале текста имеется заданная строка;
- среди всех строк сообщения, удовлетворяющих заданному шаблону, имеется строка, равная заданной.

Способ применения правил фильтрации и состав входящих в них условий определяется пользователем данного блока ВФ. Правила в необходи-

мой последовательности сохраняются в профиле пользователя.

Настройка первичной фильтрации по результатам вторичной

В данной части мы не приводим подробного описания используемых нами методов синтеза нейросети с переменной структурой и методов формирования словаря, поскольку это достаточно большая тема, требующая отдельного рассмотрения. Однако здесь мы приведем принципы формирования обучающей выборки на основании результатов голосования пользователей, которая используется для настройки нейросети.

Обучающая выборка (ОВ) реализуется в виде БД, в которой сохраняются сообщения с указанием о том, к какому типу они относятся (СНР или обычные). Формирование ОВ осуществляется по принципу *FIFO*-буфера. В обучающей выборке одновременно не может находиться более N СНР и более M обычных пользовательских сообщений. Числа N и M задаются администратором. В том случае, если число сообщений превысит одно из заданных значений, то из БД ОВ вытесняются наиболее ранние сообщения с тем же статусом, что и вновь поступившее.

Занесение сообщения в БД ОВ осуществляется в момент его первичного получения. Для каждого j-го сообщения в ОВ вычисляется два выборочных уровня уверенности (ВУУ): ВУУ в том, что данное сообщение является СНР — $G_S[M_j]$; ВУУ в том, что данное сообщение — обычное сообщение пользователя $G_H[M_j\}$. Значение ВУУ является целым числом от 0 до 100.

Сообщение M_j признается СНР, если

$$G_S[M_j] > G_H[M_j] + H,$$

где H — порог, задаваемый администратором.

Сообщение признается пользовательским, если

$$G_H[M_j] > G_S[M_j] + H.$$

В том случае, если статус сообщения остается неопределенным в течение заданного интервала времени, то оно удаляется из ОВ. В процессе обучения принимают участие только те сообщения, которые имеют определенный статус.

В том случае, если имеется уверенность в том, что данное сообщение является СНР, то $G_S[M_j] = 100$; $G_H[M_j] = 0$.

Если имеется уверенность в том, что данное сообщение обычное пользовательское, то

$$G_H[M_j] = 100; \quad G_S[M_j] = 0.$$

В противном случае значения $G_S[M_j]$ и $G_H[M_j]$ рассчитывается по методу, приводимому ниже.

Администратором определяются четыре весовых параметра W_S, где $S \in \{SA, SM, HA, HM\}$. Каждый параметр показывает значимость соответствующего статуса для величин $G_S[M_j]$ и $G_H[M_j]$. Параметры W_S должны принимать значения в пределах от 0 до 1. Чем больше W_S, тем выше вклад. Типичными значениями являются $W_{SM} = W_{HM} = 1$; $W_{SA} = W_{HA} = 0{,}5$. Определим величину квалификации i-го пользователя относительно решений администратора:

$$c_i = \frac{\sum_{M \in U_i^A} \left\lfloor (f_A^{SM}[M] + f_A^{SA}[M])(f_i^{SM}[M] + f_i^{SA}[M]) \right\rfloor + \\ + \left\lfloor (f_A^{HM}[M] + f_A^{HA}[M])(f_i^{HM}[M] + f_i^{HA}[M]) \right\rfloor}{|U_i^A|},$$

где U_i^A — множество сообщений, которые получил данный пользователь и по которым имеется решение администратора. Тогда

$$G_S[M_j] = \frac{\sum_{i \in U[M_j]} c_i \left\lfloor f_i^{SA} W_{SA} + f_i^{SM} W_{SM} \right\rfloor}{\sum_{i \in U[M_j]} c_i} 100;$$

$$G_H[M_j] = \frac{\sum_{i \in U[M_j]} c_i \left\lfloor f_i^{HA} W_{HA} + f_i^{HM} W_{HM} \right\rfloor}{\sum_{i \in U[M_j]} c_i} 100.$$

В формулах для G_S и G_H индекс i пробегает по всем пользователям, получившим сообщение M_j.

Приведенные формулы позволяют формировать выборку сообщений даже без непосредственного участия администратора. В процессе обучения величина $G_S[M_j]$ ($G_H[M_j]$) определяет частоту, с которой образ сообщения M_j появляется в обучающей последовательности. Чем выше значение, тем чаще образ применяется при обучении.

Выводы

В данной статье рассмотрен новый подход к реализации ФИП НР, в котором наряду с классическими методами ручной настройки фильтров применяются автоматические методы настройки, учитывающие результаты голосования пользователей по отдельным сообщениям. Голоса пользователей учитываются в процессе их поступления в

реальном масштабе времени. При голосовании учитывается квалификация пользователей, определяющаяся на основании совпадения их решения с решением администратора.

Такой подход позволяет повысить эффективность ФИП и обеспечить своевременную реакцию системы фильтров на возникающие ошибки ФИП.

Список литературы

1. **Cranor L. F. and LaMacchia B. A**. Spam! Communications of the ACM. 1998. Vol. 41. N 8. P. 74—83.
2. **Why** Am I Getting All This Spam? Unsolicited Commercial E-mail Research Six Month Report. Washington: The Center for Democracy and Technology, 2002. 18 p.
3. **Рыбак А.** Как не погибнуть под лавиной // Chip. Август. 2003.
4. **Atkins S.** Size and Cost of the Problem // In Proc. of the 56th Internet Engineering Task Force Meeting, San Francisco, March. 2003. 31 p.
5. **Raymond Eric S.** Bogofilter. http://bogofilter.sourceforge.net/
6. **SpamAssassin.** http://spamassassin.org/
7. **Prakash,** Vipul Ved. Vipul's Razor. http://razor.sourceforge.net/
8. **SpamTest**/Kaspersky AntiSpam, http://www.ashmanov.com
9. **Галушкин А. И**. Теория нейронных сетей. Кн. 1: Учеб. пособие для вузов. М.: ИПРЖР, 2000.