Создание ловушек для спама и сравнение антиспамов

Андрей Бондаренко
(abondarenko@gmail.com)

Источники спама

- Провайдеры
- Большие организации
- Собственные ловушки

Потоки спама очень сильно привязаны к региону, по этому судить о собственных успехах в США не имея потока спама для американцев нельзя.

Создание ловушек для спама

- Регистрация на известных форумах
- Регистрация на порно- и варез- ресурсах
- Создание видимости «живого человека»
- Регистрация на ресурсах, рассылающих легитимные новости и рекламу
- Usenet, IRC, MMORG, игры и т.п.

Создание ловушек для спама

- Продолжительность работы – 2 года
- Поток из 200-300 сообщений в день, что очень мало. Для сравнения – на личный адрес, которому около 5 лет приходит до 500 сообщений
- Параллельно 1 из адресов подписывали только на рекламу, чтобы проверять себя на ложные срабатывания

Создание ловушек для спама

- В потоке легитимной рекламы за 2 года спама стало больше чем рекламы
- Характер спама менялся, если адреса были «засвечены» на затрояненных компьютерах
- Создание нового домена – новый хороший поток спама без усилий с нашей стороны

Как отправляется спам?

- Dial-up прямой рассылкой – прошлый век
- Через open-relay – прошлый век
- Через сервер провайдера и свой сервер – прошлый век
- Через быстрый ботнет – очень хороший метод
- Через медленный ботнет – тоже хороший метод, но хуже и дороже быстрого ботнета

Выводы

- Засветка адреса в форуме малоэффективна
- Засветка адреса на затрояненном компьютере очень эффективна, но не приемлема по этическим соображениям
- Крупные и уважаемые компании иногда допускают утечки личных данных, неуважаемые ресурсы – всегда.

Определение спама

Спам это незапрошенная массовая технически анонимная рассылка.

Важность определения

- Нельзя ловить нормальную рекламу
- Нельзя опираться только на массовость
- Нельзя наказывать заказчика спама
- Наказывать «последнюю милю» доставки спама слишком строго - неправильно
- Некоторый класс хостов нужно рассматривать как рассылающий только спам, никакого SMTP для xDSL

Методы фильтрации

- Контроль доступа
- Контроль содержимого статистическими методами
- Контроль содержимого эвристическими методами

Контроль доступа

- Черные списки
- Белые списки
- Авторизация
- Серые списки

Методы очень дешевы и эффективны, но допускают ложные срабатывания, которые трудно контроллировать.

Статистические методы

- Байес и его модификации
- Применяется в Spamassassin
- Применяется во всех персональных антиспамах

Очень хорош для персонального использования, но часто требует наблюдения и вмешательства администратора. Трудно бороться с ложными срабатываниями.

Эвристические методы

- Анализ заголовков на целостность и валидность
- Анализ содержательной части на наличие характеристических терминов
- Сигнатурный анализ содержательной части

Хорошо работает при правильном подходе, но требует спам-лаборатории у вендора

Сравнение антиспамов

- Detection Rate – отношение распознанных как спам писем ко всему объему спама.
- False Positive Rate – отношение ошибочно распознанных как спам писем ко всему объему неспама.
- Реальные цифры из SLA MessageLabs:
  - DR = 95%, FPR = 0.0003%

DR и FPR нельзя посчитать роботом и графигки работы антиспама, собранные по статистике работы движка, о них тоже ничего не говорят.

Сравнение антиспамов

- Тестирование должно быть достаточно долгим
- Тестировать можно только на живых потоках, коллекции спама ничего не говорят о реальном качестве работы
- Анализ теста должен проводить живой человек
- Если тест показывает очень хороший или очень плохой результат, это значит, что его проводили некачественно

Всё неэффективно по отдельности

Менее 75% детектирования у текущих антиспамовых решений при правильном их использовании и не бывает

MessageLabs – детектирование 95% по SLA, что при 100 писем в день уже заметно раздражает пользователя.

Лучшее решение – серверный коммерческий антиспам + персональный антиспам в почтовом клиенте.

Вопросы?

Спасибо за внимание