

**ADWINS**  
интернет-проекты

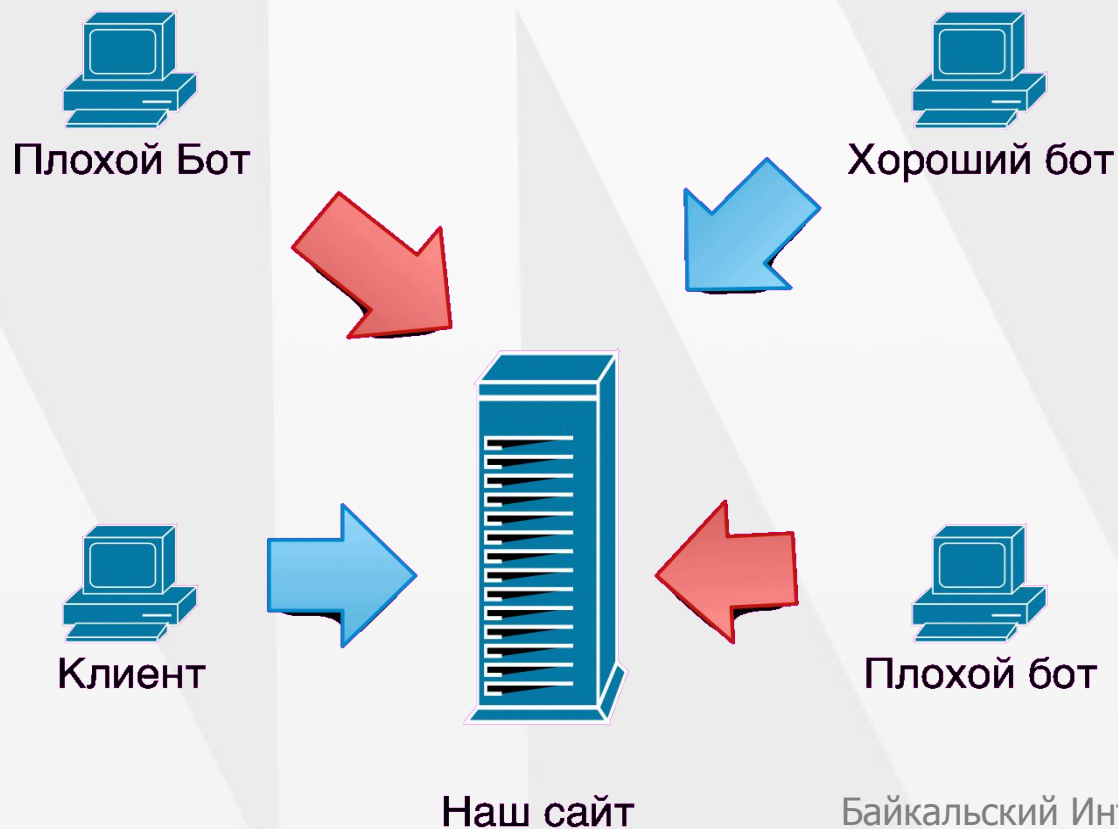


Две стороны медали:  
сбор информации с сайтов  
и защита от него

Дмитрий Сизиков  
программист

Байкальский Интернет Форум  
Иркутск, 19 августа 2011 г.

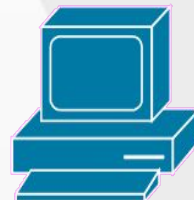
## К нашему сайту постоянно обращаются боты



## Зачем они приходят?



Плохой бот



Хороший бот

## Что в этом плохого?

- Нагрузка
- Бесконтрольное распространение вашего контента
- Замусоривание
- Проблемы с поисковыми системами
- Подрыв доверия пользователей

## Есть ли плюсы?

Две стороны медали:  
сбор информации с сайтов и защита от него

## Как можно использовать полученную информацию?

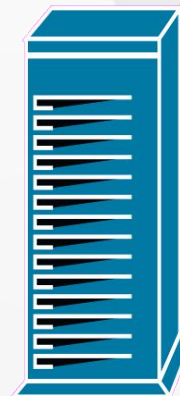
## Основные принципы защиты

- Полной защиты не существует
- Сделать сбор данных нерентабельным
- Комплексный подход к проблеме
- Набор методов защиты зависит от конкретной задачи
- Пользовательские данные прежде всего
- Не навреди клиенту



# Метод первый. Балльно-рейтинговая система

**Метод второй.  
Ограничение скорости выдачи страниц,  
подсчет времени работы со страницей**

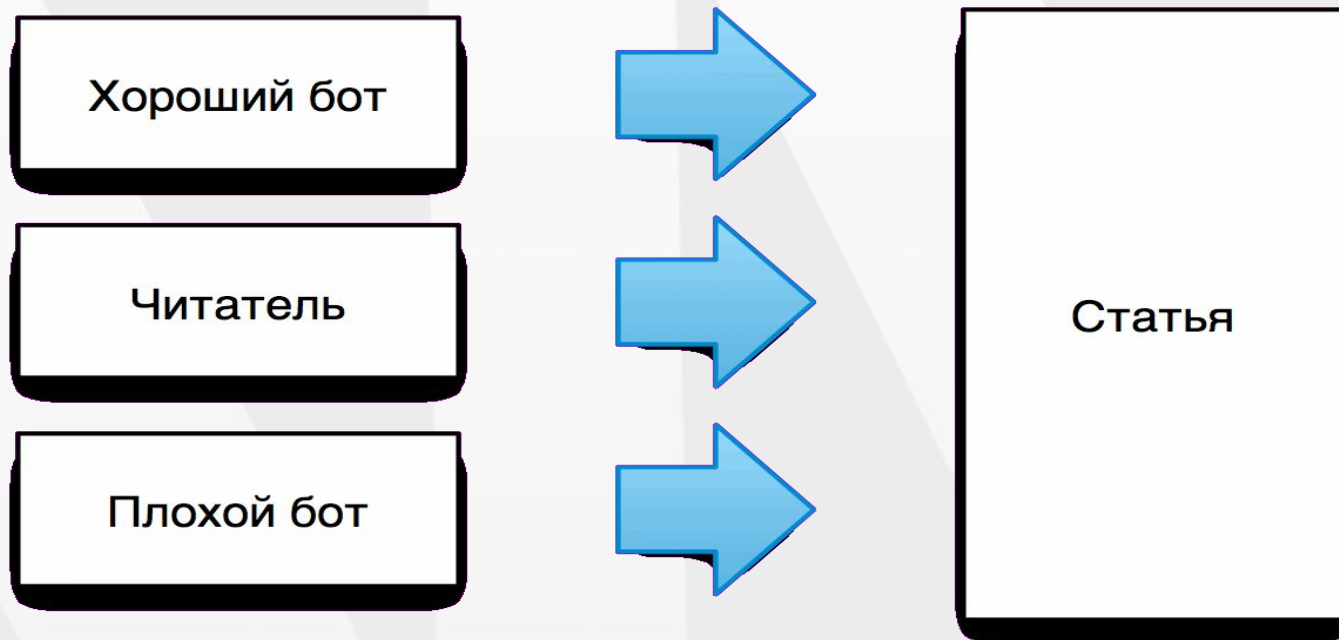


Наш сайт

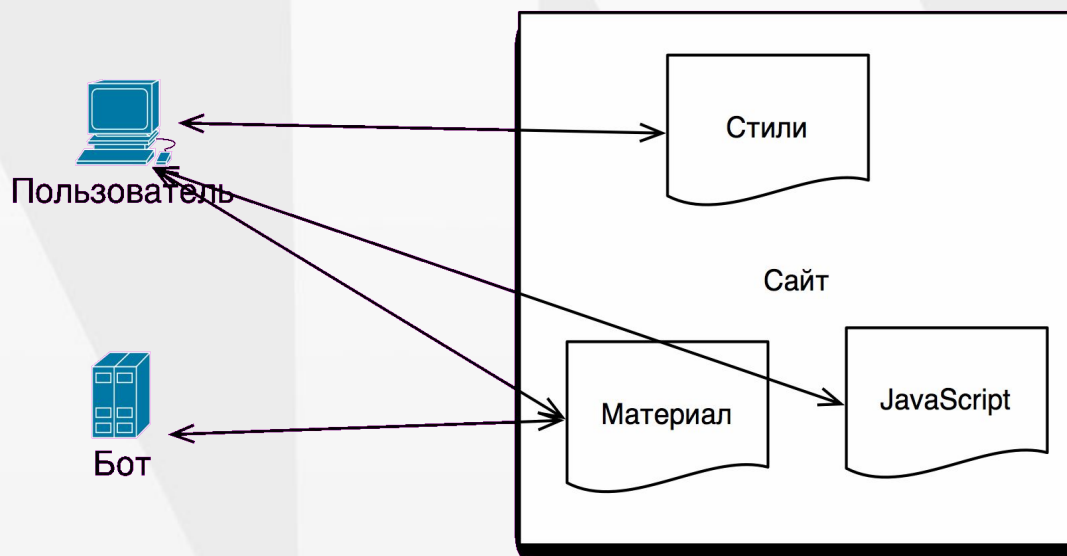
## Метод третий. Ловушки

- JavaScript-ловушки
- Визуальные ловушки
- Скрытые поля

## Метод четвертый. Материал на сайт при приходе хороших ботов



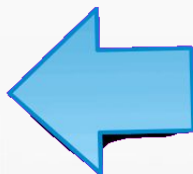
## Метод пятый. Анализ загрузки стилей и картинок



## Метод шестой. Замусоривание контента



Бот



Фраза из вашей статьи

Фраза 33df из sdfff вашей  
bla-bla статьи

## Метод седьмой. Анализ действий пользователя на сайте

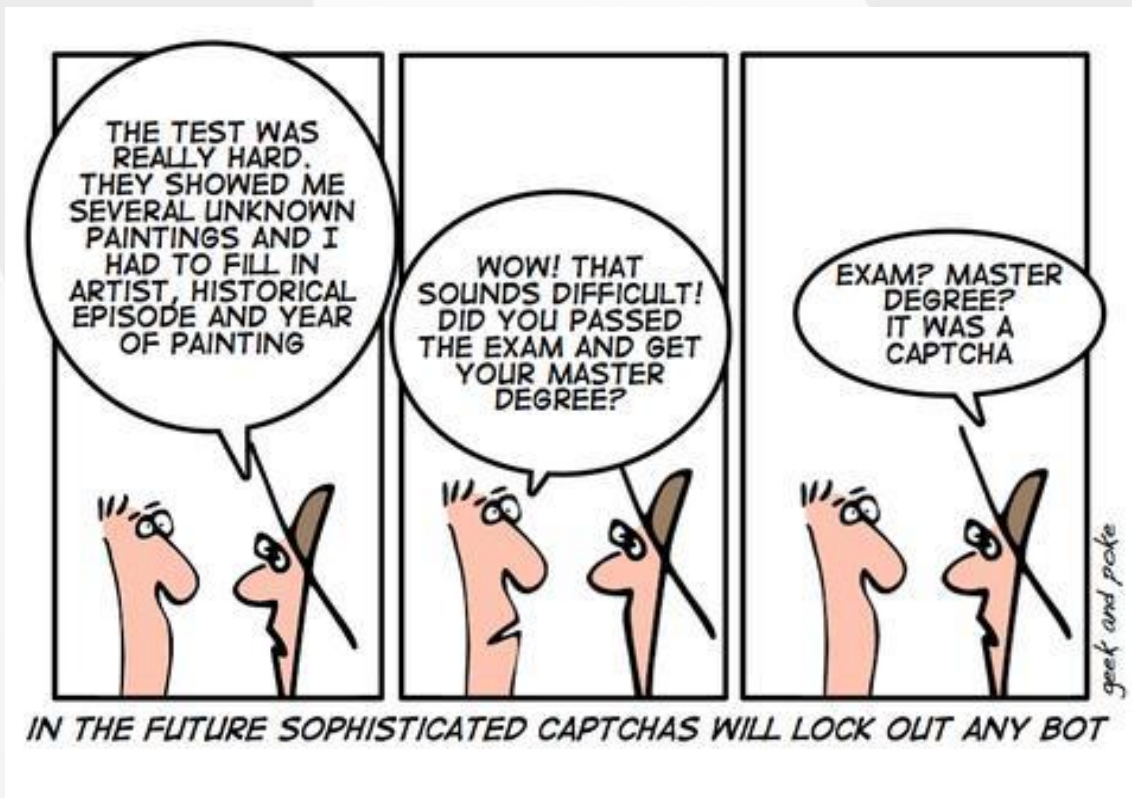
- Клики
- Движение мышкой
- и т. д.

## Метод восьмой. Анализ контента, генерируемого пользователем

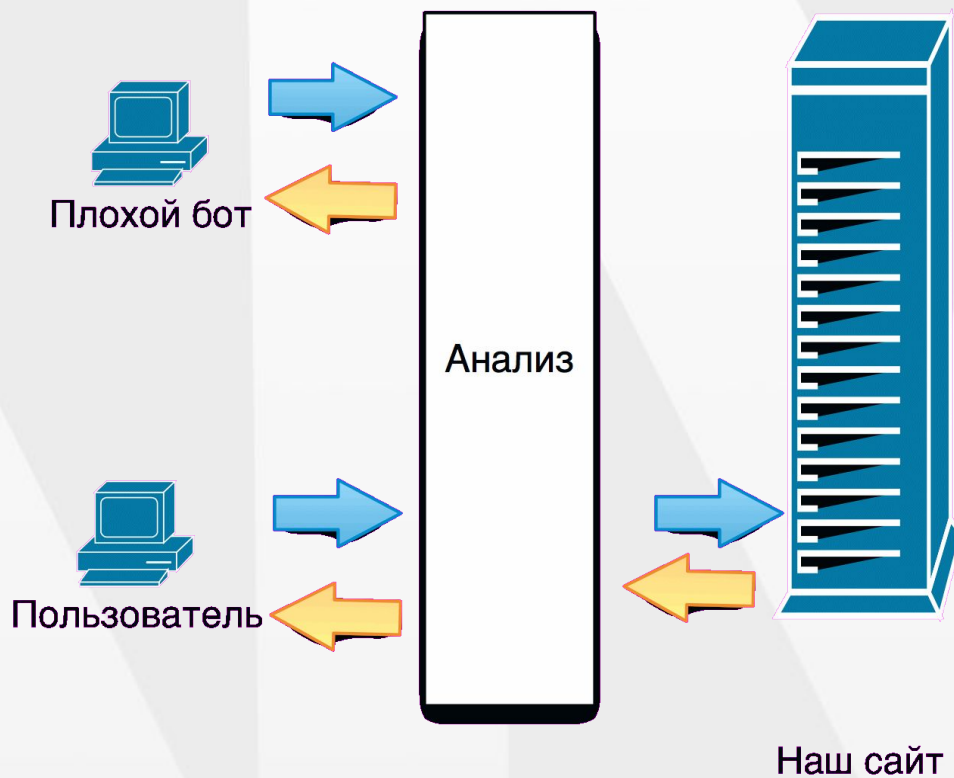
- Однотипные комментарии
- Множество ссылок
- Карма пользователя
- И т. д.



## Метод девятый. Captcha



## Метод десятый. Обман бота



**100% метод защиты вашего контента:**

**Не показывайте его никому :]**

Спасибо за внимание!

Вопросы?

**Дмитрий Сизиков**  
программист

[suzrik@boardev.com](mailto:suzrik@boardev.com)    [www.adwins.ru](http://www.adwins.ru)