登录

黑帽SEO剖析之总结篇

nmask / 2017-09-28 09:35:00 / 浏览数 3694 技术文章 技术文章 顶(0) 踩(0)

此系统文章总共分为四篇，分别是手法篇、工具篇、隐藏篇、总结篇；本篇为总结篇，主要介绍黑帽seo行为的检测以及预防。可以说此系列前面三篇文章，是为最后一篇做

如何检测自身网站是否被劫持？

前面介绍了很多关于黑帽seo的手法，那作为站长或者运维该怎么去监控自身网站是否被入侵，且被黑帽seo利用了呢？这里不说如何检测入侵，因为这不是本文的范畴，我

内部监控

可以监控服务器web目录下的文件改动情况，一般黑帽seo都需要改动web目录下的文件（新增文件，或更改文件内容）。当然有些只改变nginx配置就可以达到目的，因此

小结：内部监控比较类似防篡改的检测，只是面对网页劫持，除了响应文件内容改动以外，还需要响应新增文件等行为，包括服务器配置文件的改动。

外部检测

黑帽seo手法从根本上是欺骗搜索引擎，因此检测本质上也可以从搜索引擎出发。检测网站在搜索引擎搜索显示下是否出现了敏感的内容，比如：博彩、色情等。由于网页劫

多维度包括但不局限于以下几种：

- 采用不同地区的IP检测目标网站
- 采用不同时间段内检测目标网站
- 采用不同的UA访问目标网站
- 采用不同的访问方式目标网站（百度搜索跳转、直接访问域名）

检测步骤分为：

- 获取搜索引擎搜索结果
- 模拟浏览器访问搜索结果网页
- 解析网页源码等元素
- 匹配规则判断网站是否被劫持

获取搜索引擎搜索结果

这一步骤需要爬取搜索引擎，比如我们要判断thief.one网站是否被劫持，可以搜索百度：site:thief.one
色情。关键词需要自己搜集，然后利用爬虫爬取百度的搜索结果。
显然这一步需要对抗百度搜索引擎，防止被其屏蔽问题，还要能够正确的获取百度的搜索结果。关于爬起搜索引擎可参考：
爬取搜索引擎之寻你千百度
爬取搜索引擎之搜狗

模拟浏览器访问搜索结果网页

当爬到所需要的网页链接后，我们需要重放url获取信息。这一步需要能够动态执行网页中嵌入的js代码，动态跟踪网页的走向（跳转）。这里推荐使用phantomjs当然也可

解析网页源码等元素

可以利用python解析网页源码、网页标题、URL、js等内容，最方便的做法是获取各个参数的内容，处理数据打标后扔到机器学习的算法中进行模型计算。

匹配规则判断网站是否被劫持

可以使用正则等方式，根据黑帽seo等特征建立规则库去匹配。当然也可以利用机器学习的方式去对相关网页进行分类，我们曾经使用过某种算法，将准确率提高到了90%左

小结：外部检测难度比较大，目前黑帽seo主要针对百度，因此这相当于去检测百度的搜索结果；而如何模拟浏览器访问也是一大难题，当然最重要的是最后的机器学习，如

谁来为此买单？

基于黑帽SEO大多数都为博彩赌博行业做推广，将会增加网民沉迷网络赌博的风险，纵观身边因为网络赌博而家破人亡的事情不在少数；而也有一部分黑帽SEO在为枪支弹药
首先网站管理者难辞其咎，正因为管理员安全意识的淡薄，网站安全性不高，导致被入侵最终成为黑产的一部分。在我自身处理的几起类似事件中，网站管理员往往是一副无
其次搜索引擎应该担负一定的责任，因为黑帽SEO行为主要针对搜索引擎，说白了就是利用搜索引擎算法漏洞，提升非法网站权重。国内大多数网民上网都使用搜索引擎。搜

如何制止与防御？

如果您是网民，制止黑帽seo最好的方式就是科学上网，发现非法网站及时提交到安全联盟或向搜索引擎举报。
如果您是网站管理员，请做好自身网站的安全建设，及时补漏；若已发现被入侵，及时联系技术人员处理。

谈谈心

当在写这篇文章前，我思索着尽量能够全面地介绍黑帽SEO知识以及手法。当开始写这篇文章的时候，我便有点无从下手，因为涉及知识面太广，手法又非常丰富，我研究黑

传送门

黑帽SEO剖析之总结篇
黑帽SEO剖析之隐身篇
黑帽SEO剖析之工具篇
黑帽SEO剖析之手法篇

上一篇：黑帽SEO剖析之隐身篇 下一篇：黑帽SEO剖析之隐身篇

1. 0 条回复
   - 动动手指，沙发就是你的了！