需求二

**针对各模块采取相应的分析算法，重点分析在页面的敏感请求接口处（登陆、注册等）产生的日志。多维度、全面的分析不同类型的攻击数据的特征，提取特征字段用于后续的深度学习模块。**

需求2.1：多维度分析不同类型的攻击方式

针对各种攻击方式的特点和原理，多维度、全面地分析其数据特征，选择特征字段作为恶意攻击识别模型的维度，用于后续的深度学习模块。

1. **各种攻击方式的特点和原理的分析**

**1、DDOS:** 攻击者快速的向同一个受害者发送虚假的SYN报文，那么受害者机器上就会产生大量的TCP/IP半连接，并且都要等待这些半连接超时，从而使系统资源耗尽，不能及时响应正常的服务请求。达成攻击就是因为攻击者伪造了自己的IP，使得“第二次报文”的SYN/ACK报文不能到达攻击者，同时攻击者也不会发ACK报文进行“第三次报文”，但是受害者却一直在苦苦等待“第三次”报文ACK的到来。所以，反应在日志上就是http请求长度与正常日志存在明显区别，明显**少于正常的http请求所发送的数据长度**。同时**伪造的IP**也在一定程度上暴露了攻击的目的。

**2、慢连接**：客户端连接一个主机，服务器接收连接，客户端请求一个文件，服务器发送一个应答，它的原理就是让服务器等待先于目标服务器建立连接，然后服务器再保持连接等待的时候就消耗了资源（在http请求报文当中，它的后面存在一个空行，包括回车符和换形符。换行，他的意思是告诉服务器在请求头部结束，后面就不再有请求头。如果服务器没有收到这个回车符或者换行符，就一直在那等着。然后攻击者就利用这点，发送一个http头部，一直不停的发，然后服务器就在那等啊等，消耗了服务器的资源）。因此，慢链接请求的**请求时间**显然是**很长**的。在日志上的表现就是**request\_time字段尤其长**，普遍大于50秒。同时接受方一直处于等待状态也导致了**http的请求状态**有所不同。

**3、撞库：**存在撞库的根本原因是很多互联网用户在不同网站使用的是相同的账号密码，因此攻击者可以通过获取用户在A网站的账户从而尝试登录B网站。高水准的撞库攻击不易发现，实现起来需很高的技术能力，因此成本较高，当前多数的撞库还是以单脚本登录验证、分布式脚本登录验证，自动代理登录验证。所以，黑客的恶意撞库其源头还是肉鸡、代理、botnet。通过自动识别异常IP，因为如果机器扫描的话，肯定会有一些特征能被你抓取到，比如：单位时间内操作次数，比如1分钟尝试登录超过100次、正常用户不可能有这么大的频度的；1分钟内尝试登录错误账号超过一定次数；其它的自己扩展。对于异常的IP，整理一个非常严格的库，甚至直接**禁止这些IP**访问网站。

**4、爆破：**暴力破解攻击是通过巨大的尝试次数获得一定成功率的的。因此在web（应用程序）日志上，你会经常发现有很多的登录失败条目，而且这些条目的IP地址通常还是同个IP地址。有时你又会发现不同的IP地址会使用同一个账户、不同的密码进行登录。大量的暴力破解请求会导致服务器日志中出现大量异常记录，从中你会发现一些奇怪的进站前链接（referring urls），比如：http://user:password@website.com/login.html。有时，攻击者会用不同的用户名和密码频繁的进行登录尝试，这就给检测系统或者记录关联系统一个检测到他们入侵的好机会。

综上：通过以上几种常见的互联网攻击方式可以发现，虽然攻击方式很多，但是反应到日志上大致包括以下几种特征：

1. **同一IP访问；**
2. **极短时间内大量访问；**
3. **访问时间过长；**
4. **http请求体异常；**
5. **网页状态码异常；**
6. **cookies值异常；**
7. **请求是否完成**等。

结合需求，同时加入IP+cookies等的分析方法来对可能存在的恶意访问进行分析。各种攻击方式的参考数据如下图所示：

