(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107547547 B
(45)授权公告日 2020.06.02

(21)申请号 201710792548.8

(22)申请日 2017.09.05

(65)同一申请的已公布的文献号
申请公布号 CN 107547547 A

(43)申请公布日 2018.01.05

(73)专利权人 成都知道创宇信息技术有限公司
地址 610000 四川省成都市高新区天府三街219号2栋11楼

(72)发明人 陈海洋 叶兴 张文宇 吴文林 郑斌

(74)专利代理机构 成都禾创知家知识产权代理有限公司 51284
代理人 裴娟

(51)Int.Cl.
H04L 29/06(2006.01)

(56)对比文件
CN 103473507 A,2013.12.25,
CN 106909841 A,2017.06.30,

审查员 丁彬

权利要求书1页 说明书3页

(54)发明名称

一种基于编辑距离的TCP CC识别方法

(57)摘要

本发明公开了一种基于编辑距离的TCP CC识别方法，在恶意连接到达服务器时，直接分析其发送的数据内容，因其为随机数据，必然与特征字节串不匹配，从而立即识别出恶意连接，而不是放行，当同一个恶意攻击者(IP)的连接次数达到阈值后，直接封锁该攻击者(IP)，在整个过程中，服务器上层业务不会收到任何随机数据(垃圾数据)，因而不会受到任何影响。本发明方法能够精准地识别攻击，降低误报率、漏报率，同时因其直接参与连接的过程，可以攻击较早的阶段即可识别阻断，无需等待连接结束后进行分析，极大地提高了识别的效率。

CN 107547547 B

1.一种基于编辑距离的TCP CC识别方法,其特征在于,包括以下步骤:

步骤1:设定初始的特征字节串S以及特征字节串长度为M,设定编辑距离的阈值为D,且M ≥ D ≥1;

步骤2:新连接到来时,接收前M个字节的数据,存放到缓冲区B中;

步骤3:计算特征字节串与缓冲区B中存放的内容的编辑距离d;

步骤4:比较编辑距离d与阈值D;

步骤5:若编辑距离d ≥阈值D,进行如下操作:

1)查找当前连接的源 IP 地址是否被缓存,若未被缓存则进行步骤2),否则进行步骤3);

2)缓存源 IP 地址,并记录其对应的识别失败计数为1,记录该源 IP 地址缓存超时时间为T;

3)判断该 源IP 地址的缓存超时时间T是否已超时,若超时则进行步骤4),否则进行步骤5);

4)设置该源 IP 地址的缓存记录中的识别失败次数为1,并重置该缓存超时时间为T;

5)更新源IP地址缓存记录的识别失败计数加1;

6)判断该源 IP 地址缓存记录的识别失败次数是否超过配置的阈值,若超过则进行步骤7),否则关闭该连接,结束本次识别处理;

7)关闭该连接,删除该源IP地址的缓存记录,并将该源IP地址加入系统黑名单,阻止其连接请求;

步骤6:若编辑距离d <阈值D,进行如下操作加强特征:

a)判断是否需要动态更新特征字节串,若是则进行步骤b),否则结束本次处理;

b)将提取的前M个字节与特征字节串中的字节一一对应,得到M对数据(M_i,S_i),i=1,2,3…M;

c)遍历M对数据,计算每对数据的均值$(M_i + S_i)/2$;

d)将特征字节串的内容更新为步骤c)得到的均值序列,即:$S_1' = (S_1 + M_1)/2$,$S_2' = (S_2 + M_2)/2$,…,$S_M' = (S_M + M_M)/2$,其中:S_1,S_2,…,S_M依次代表特征字节串的第一个字节,第二个字节,…,第M个字节。

一种基于编辑距离的TCP CC识别方法

技术领域

[0001] 本发明涉及CC攻击识别领域,特别是一种基于编辑距离的TCP CC识别方法。

背景技术

[0002] CC是指攻击者借助代理服务器生成指向受害主机的合法请求,实现DOS和伪装。CC攻击主要针对WEB应用程序比较消耗资源的地方进行疯狂请求,比如,论坛中的搜索功能,如果不加以限制,任由人搜索,普通配置的服务器在几百个并发请求下,MYSQL服务就会瘫痪。CC攻击的种类有三个,分别为直接攻击、代理攻击和僵尸网络攻击。

[0003] 防御CC攻击可以通过多种方法,禁止网站代理访问,尽量将网站做成静态页面,限制连接数量,修改最大超时时间等。传统的TCP CC攻击识别主要通过计算单位时间请求数量或者单个IP的请求频率是否达到阈值来识别。其存在以下不足:1)如果将阈值设定较高,会有很多低频攻击无法识别;2)如果将阈值设定较低,会有很多正常请求被错误识别为攻击;3)攻击识别效率较低,需要在攻击已经产生后才开始识别攻击,并需要持续一段时间才能完成识别。

[0004] CC(Challenge Collapsar)攻击:DDoS攻击的一种,通过伪装合法请求进行攻击;

[0005] 编辑距离(Edit Distance):指两个字符串之间,由一个转成另一个所需的最少编辑操作次数。

发明内容

[0006] 本发明所要解决的技术问题是提供一种基于编辑距离的TCP CC识别方法,对于相同的业务来说,其TCP连接的前N个字节基本是固定或者相似的,本方法通过提取TCP第一个数据包的内容,并将该内容与历史学习到的内容进行编辑距离的计算,依据计算结果识别TCP的CC攻击,提高识别的效率。

[0007] 为解决上述技术问题,本发明采用的技术方案是:

[0008] 一种基于编辑距离的TCP CC识别方法,包括以下步骤:

[0009] 步骤1:设定初始的特征字节串S以及特征字节串长度为M,设定编辑距离的阈值为D,且M ≥ D ≥1;

[0010] 步骤2:新连接到来时,接收前M个字节的数据,存放到缓冲区B中;

[0011] 步骤3:计算特征字节串与缓冲区B中存放的内容的编辑距离d;

[0012] 步骤4:比较编辑距离d与阈值D;

[0013] 步骤5:若编辑距离d ≥阈值D,进行如下操作:

[0014] 1)查找当前连接的源 IP 地址是否被缓存,若未被缓存则进行步骤2),否则进行步骤3);

[0015] 2)缓存源 IP 地址,并记录其对应的识别失败计数为1,记录该源 IP 地址 缓存超时时间为T;

[0016] 3)判断该源 IP 地址的缓存超时时间T是否已超时,若超时则进行步骤4),否则进

行步骤5)；

[0017] 4)设置该源 IP 地址的缓存记录中的识别失败次数为1,并重置该记录缓存超时时间为T；

[0018] 5)更新源IP地址缓存记录的识别失败计数加1；

[0019] 6)判断该 源IP 地址缓存记录的识别失败次数是否超过配置的阈值,若超过则进行步骤7),否则关闭该连接,结束本次识别处理；

[0020] 7)关闭该连接,删除该源IP地址的缓存记录,并将该源IP地址加入系统黑名单,阻止其连接请求；

[0021] 步骤6:若编辑距离d <阈值D,进行如下操作加强特征：

[0022] a)判断是否需要动态更新特征字节串,若是则进行步骤b),否则结束本次处理；

[0023] b)将提取的前M个字节与特征字节串中的字节一一对应,得到M对数据(M_i,S_i),i=1,2,3…M；

[0024] c)遍历M对数据,计算每对数据的均值(M_i + S_i)/2；

[0025] d)将特征字节串的内容更新为步骤c)得到的均值序列,即:S_1'=(S_1 + M_1)/2,S_2'=(S_2 + M_2)/2,…,S_M'=(S_M + M_M)/2,其中:S1,S2,…,S_M依次代表特征字节串的第一个字节,第二个字节,…,第M个字节。

[0026] 与现有技术相比,本发明的有益效果是:当攻击者恶意连接服务器后,发送随机数据(垃圾数据),此时使用传统的识别防御方法,需要判断攻击者的IP发起的连接数量,并需要在一段时间内对恶意连接放行,以等待其达到阈值。使用本发明后,在恶意连接到达服务器时,直接分析其发送的数据内容,因其为随机数据,必然与特征字节串不匹配,从而立即识别出恶意连接,而不是放行,当同一个恶意攻击者(IP)的连接次数达到阈值后,直接封锁该攻击者(IP),在整个过程中,服务器上层业务不会收到任何随机数据(垃圾数据),因而不会受到任何影响。

[0027] 通过比对报文内容,本发明方法能够精准地识别攻击,降低误报率、漏报率,同时因其直接参与连接的过程,可以攻击较早的阶段即可识别阻断,无需等待连接结束后进行分析,极大地提高了识别的效率。

具体实施方式

[0028] 下面通过具体实施方式对本发明技术方案做详细的说明。

[0029] 1、设定初始的特征字节串S以及特征字节串长度为M,设定编辑距离的阈值为D(M ≥ D ≥1)；

[0030] 2、新连接到来时,接收前M个字节的数据,存放到缓冲区B中；

[0031] 3、计算特征字节串与缓冲区B中存放的内容的编辑距离d；

[0032] 4、比较 d 与 D 的值；

[0033] 5、若 d ≥ D,进行如下操作：

[0034] 1)查找当前连接的源 IP 地址是否被缓存,若未被缓存则进行步骤2),否则进行步骤3)；

[0035] 2)缓存源 IP 地址,并记录其对应的识别失败计数为1,记录该源 IP 地址缓存超时时间为T(T可配置)；

[0036] 3)判断该 源IP 地址的缓存超时时间T是否已超时,若超时则进行步骤4),否则进行步骤5);

[0037] 4)设置该源 IP 地址的缓存记录中的识别失败次数为1,并重置该缓存超时时间为T;

[0038] 5)更新源IP地址缓存记录的识别失败计数加1;

[0039] 6)判断该源 IP 地址缓存记录的识别失败次数是否超过配置的阈值,若超过则进行步骤7),否则关闭该连接,结束本次识别处理;

[0040] 7)关闭该连接,删除该源IP地址的缓存记录,并将该源IP地址加入系统黑名单,阻止其连接请求;

[0041] 6、若编辑距离d <阈值D,进行如下操作加强特征:

[0042] a、判断是否需要动态更新特征字节串,若是则进行步骤b),否则结束本次处理;

[0043] b、将提取的前M个字节与特征字节串中的字节一一对应,得到M对数据(M_i,S_i);

[0044] c、遍历M对数据,计算每对数据的均值$(M_i + S_i)/2$;

[0045] d、将特征字节串的内容更新为步骤c得到的均值序列,即:$S_1' = (S_1 + M_1)/2$,$S_2' = (S_2 + M_2)/2$,…,$S_M' = (S_M + M_M)/2$。

[0046] 由编辑距离引申出的字符串相似度,可以用作二进制字节的相似度替换本发明中的编辑距离方法。