1. **TCP连接基本特征（共9种）**

基本连接特征包含了一些连接的基本属性，如连续时间，协议类型，传送的字节数等。

一个网络连接定义为在某个时间内从开始到结束的TCP数据包序列，并且在这段时间内，数据在预定义的协议下（如TCP、UDP）从源IP地址到目的IP地址的传递。每个网络连接被标记为正常（normal）或异常（attack），异常类型被细分为4大类共39种攻击类型，其中22种攻击类型出现在训练集中，另有17种未知攻击类型出现在测试集中。

（1）duration. 连接持续时间，以秒为单位，连续类型。范围是 [0, 58329] 。它的定义是从TCP连接以3次握手建立算起，到FIN/ACK连接结束为止的时间；若为UDP协议类型，则将每个UDP数据包作为一条连接。数据集中出现大量的duration = 0 的情况，是因为该条连接的持续时间不足1秒。

（2）protocol\_type. 协议类型，离散类型，共有3种：TCP, UDP, ICMP。

（3）service. 目标主机的网络服务类型，离散类型，共有70种。’aol’, ‘auth’, ‘bgp’, ‘courier’, ‘csnet\_ns’, ‘ctf’, ‘daytime’, ‘discard’, ‘domain’, ‘domain\_u’, ‘echo’, ‘eco\_i’, ‘ecr\_i’, ‘efs’, ‘exec’, ‘finger’, ‘ftp’, ‘ftp\_data’, ‘gopher’, ‘harvest’, ‘hostnames’, ‘http’, ‘http\_2784′, ‘http\_443′, ‘http\_8001′, ‘imap4′, ‘IRC’, ‘iso\_tsap’, ‘klogin’, ‘kshell’, ‘ldap’, ‘link’, ‘login’, ‘mtp’, ‘name’, ‘netbios\_dgm’, ‘netbios\_ns’, ‘netbios\_ssn’, ‘netstat’, ‘nnsp’, ‘nntp’, ‘ntp\_u’, ‘other’, ‘pm\_dump’, ‘pop\_2′, ‘pop\_3′, ‘printer’, ‘private’, ‘red\_i’, ‘remote\_job’, ‘rje’, ‘shell’, ‘smtp’, ‘sql\_net’, ‘ssh’, ‘sunrpc’, ‘supdup’, ‘systat’, ‘telnet’, ‘tftp\_u’, ‘tim\_i’, ‘time’, ‘urh\_i’, ‘urp\_i’, ‘uucp’, ‘uucp\_path’, ‘vmnet’, ‘whois’, ‘X11′, ‘Z39\_50′。

（4）flag. 连接正常或错误的状态，离散类型，共11种。’OTH’, ‘REJ’, ‘RSTO’, ‘RSTOS0′, ‘RSTR’, ‘S0′, ‘S1′, ‘S2′, ‘S3′, ‘SF’, ‘SH’。它表示该连接是否按照协议要求开始或完成。例如SF表示连接正常建立并终止；S0表示只接到了SYN请求数据包，而没有后面的SYN/ACK。其中SF表示正常，其他10种都是error。

（5）src\_bytes. 从源主机到目标主机的数据的字节数，连续类型，范围是 [0, 1379963888]。

（6）dst\_bytes. 从目标主机到源主机的数据的字节数，连续类型，范围是 [0. 1309937401]。

（7）land. 若连接来自/送达同一个主机/端口则为1，否则为0，离散类型，0或1。

（8）wrong\_fragment. 错误分段的数量，连续类型，范围是 [0, 3]。

（9）urgent. 加急包的个数，连续类型，范围是[0, 14]。

**2. TCP连接的内容特征（共13种）**

对于U2R和R2L之类的攻击，由于它们不像DoS攻击那样在数据记录中具有频繁序列模式，而一般都是嵌入在数据包的数据负载里面，单一的数据包和正常连接没有什么区别。为了检测这类攻击，Wenke Lee等从数据内容里面抽取了部分可能反映入侵行为的内容特征，如登录失败的次数等。

（10）hot. 访问系统敏感文件和目录的次数，连续，范围是 [0, 101]。例如访问系统目录，建立或执行程序等。

（11）num\_failed\_logins. 登录尝试失败的次数。连续，[0, 5]。

（12）logged\_in. 成功登录则为1，否则为0，离散，0或1。

（13）num\_compromised. compromised条件（\*\*）出现的次数，连续，[0, 7479]。

（14）root\_shell. 若获得root shell 则为1，否则为0，离散，0或1。root\_shell是指获得超级用户权限。

（15）su\_attempted. 若出现”su root” 命令则为1，否则为0，离散，0或1。

（16）num\_root. root用户访问次数，连续，[0, 7468]。

（17）num\_file\_creations. 文件创建操作的次数，连续，[0, 100]。

（18）num\_shells. 使用shell命令的次数，连续，[0, 5]。

（19）num\_access\_files. 访问控制文件的次数，连续，[0, 9]。例如对 /etc/passwd 或 .rhosts 文件的访问。

（20）num\_outbound\_cmds. 一个FTP会话中出站连接的次数，连续，0。数据集中这一特征出现次数为0。

（21）is\_hot\_login.登录是否属于“hot”列表（\*\*\*），是为1，否则为0，离散，0或1。例如超级用户或管理员登录。

（22）is\_guest\_login. 若是guest 登录则为1，否则为0，离散，0或1。

**3. 基于时间的网络流量统计特征 （共9种，23～31）**

由于网络攻击事件在时间上有很强的关联性，因此统计出当前连接记录与之前一段时间内的连接记录之间存在的某些联系，可以更好的反映连接之间的关系。这类特征又分为两种集合：一个是 “same host”特征，只观察在过去两秒内与当前连接有相同目标主机的连接，例如相同的连接数，在这些相同连接与当前连接有相同的服务的连接等等；另一个是 “same service”特征，只观察过去两秒内与当前连接有相同服务的连接，例如这样的连接有多少个，其中有多少出现SYN错误或者REJ错误。

（23）count. 过去两秒内，与当前连接具有相同的目标主机的连接数，连续，[0, 511]。

（24）srv\_count. 过去两秒内，与当前连接具有相同服务的连接数，连续，[0, 511]。

（25）serror\_rate. 过去两秒内，在与当前连接具有相同目标主机的连接中，出现“SYN” 错误的连接的百分比，连续，[0.00, 1.00]。

（26）srv\_serror\_rate. 过去两秒内，在与当前连接具有相同服务的连接中，出现“SYN” 错误的连接的百分比，连续，[0.00, 1.00]。

（27）rerror\_rate. 过去两秒内，在与当前连接具有相同目标主机的连接中，出现“REJ” 错误的连接的百分比，连续，[0.00, 1.00]。

（28）srv\_rerror\_rate. 过去两秒内，在与当前连接具有相同服务的连接中，出现“REJ” 错误的连接的百分比，连续，[0.00, 1.00]。

（29）same\_srv\_rate. 过去两秒内，在与当前连接具有相同目标主机的连接中，与当前连接具有相同服务的连接的百分比，连续，[0.00, 1.00]。

（30）diff\_srv\_rate. 过去两秒内，在与当前连接具有相同目标主机的连接中，与当前连接具有不同服务的连接的百分比，连续，[0.00, 1.00]。

（31）srv\_diff\_host\_rate. 过去两秒内，在与当前连接具有相同服务的连接中，与当前连接具有不同目标主机的连接的百分比，连续，[0.00, 1.00]。

注：这一大类特征中，23、25、27、29、30这5个特征是 “same host” 特征，前提都是与当前连接具有相同目标主机的连接；24、26、28、31这4个特征是 “same service” 特征，前提都是与当前连接具有相同服务的连接。

**4. 基于主机的网络流量统计特征 （共10种，32～41）**

基于时间的流量统计只是在过去两秒的范围内统计与当前连接之间的关系，而在实际入侵中，有些 Probing攻击使用慢速攻击模式来扫描主机或端口，当它们扫描的频率大于2秒的时候，基于时间的统计方法就无法从数据中找到关联。所以Wenke Lee等按照目标主机进行分类，使用一个具有100个连接的时间窗，统计当前连接之前100个连接记录中与当前连接具有相同目标主机的统计信息。

（32）dst\_host\_count. 前100个连接中，与当前连接具有相同目标主机的连接数，连续，[0, 255]。

（33）dst\_host\_srv\_count. 前100个连接中，与当前连接具有相同目标主机相同服务的连接数，连续，[0, 255]。

（34）dst\_host\_same\_srv\_rate. 前100个连接中，与当前连接具有相同目标主机相同服务的连接所占的百分比，连续，[0.00, 1.00]。

（35）dst\_host\_diff\_srv\_rate. 前100个连接中，与当前连接具有相同目标主机不同服务的连接所占的百分比，连续，[0.00, 1.00]。

（36）dst\_host\_same\_src\_port\_rate. 前100个连接中，与当前连接具有相同目标主机相同源端口的连接所占的百分比，连续，[0.00, 1.00]。

（37）dst\_host\_srv\_diff\_host\_rate. 前100个连接中，与当前连接具有相同目标主机相同服务的连接中，与当前连接具有不同源主机的连接所占的百分比，连续，[0.00, 1.00]。

（38）dst\_host\_serror\_rate. 前100个连接中，与当前连接具有相同目标主机的连接中，出现SYN错误的连接所占的百分比，连续，[0.00, 1.00]。

（39）dst\_host\_srv\_serror\_rate. 前100个连接中，与当前连接具有相同目标主机相同服务的连接中，出现SYN错误的连接所占的百分比，连续，[0.00, 1.00]。

（40）dst\_host\_rerror\_rate. 前100个连接中，与当前连接具有相同目标主机的连接中，出现REJ错误的连接所占的百分比，连续，[0.00, 1.00]。

（41）dst\_host\_srv\_rerror\_rate. 前100个连接中，与当前连接具有相同目标主机相同服务的连接中，出现REJ错误的连接所占的百分比，连续，[0.00, 1.00]。

5、特征类型

duration,protocol\_type,service,flag,src\_bytes,dst\_bytes,land,wrong\_fragment,urgent,count,srv\_count,serror\_rate,srv\_serror\_rate,rerror\_rate,srv\_rerror\_rate,same\_srv\_rate,diff\_srv\_rate,srv\_diff\_host\_rate,dst\_host\_count,dst\_host\_srv\_count,dst\_host\_same\_srv\_rate,dst\_host\_diff\_srv\_rate,dst\_host\_same\_src\_port\_rate,dst\_host\_srv\_diff\_host\_rate,dst\_host\_serror\_rate,dst\_host\_srv\_serror\_rate,dst\_host\_rerror\_rate,dst\_host\_srv\_rerror\_rate