

权利要求书

1. 一种 URL 攻击检测方法，所述方法包括：
从 URL 访问请求中携带的信息中提取若干维度的特征；
将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到
5 所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于
孤立森林 Isolation Forest 机器学习算法训练得到的机器学习模型；
基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。
2. 根据权利要求 1 所述的方法，所述方法还包括：
从若干 URL 访问请求样本携带的信息中分别提取若干维度的特征；
10 其中，所述若干 URL 访问请求样本均未被标记样本标签；
基于提取到的特征构建若干训练样本；
基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得
到所述 URL 攻击检测模型。
3. 根据权利要求 2 所述的方法，所述 URL 攻击检测模型包括基于
15 Isolation Forest 机器学习算法训练得到的 M 棵随机二叉树；
所述基于 Isolation Forest 机器学习算法对所述若干训练样本进行训
练得到所述 URL 攻击检测模型，包括：
基于从所述若干训练样本中均匀抽样出的训练样本构建出 M 个训
练样本子集；
20 从所述若干维度的特征中为各训练样本子集随机选择一分类特征作
为根节点，以及在所述分类特征的最大取值和最小取值构成的取值区间
中，为各训练样本子集随机选取一分类临界值；
将各训练样本子集中所述分类特征的取值大于等于所述分类临界值
的训练样本，和所述分类特征的取值小于所述分类临界值的训练样本，
25 分别分类为所述根节点的叶节点；以及，
将各叶节点中的训练样本作为新的训练样本子集，迭代执行以上分
类过程，直到得到的各叶节点中的训练样本不可再分类时停止。

4. 根据权利要求3所述的方法, 所述将提取到的特征输入预设的 URL 攻击检测模型进行预测计算, 得到所述 URL 访问请求的风险评分, 包括:

基于提取到的特征构建预测样本;

- 5 基于所述预测样本中各特征的取值, 从根节点开始遍历各棵随机二叉树查找与所述预测样本对应的叶节点;

计算查找到的叶节点在各棵随机二叉树中的路径深度的平均值, 并对所述平均值进行归一化处理, 得到所述 URL 访问请求的风险评分。

5. 根据权利要求1所述的方法, 所述信息包括: 域名信息, 和/或
10 URL 参数; 所述若干维度的特征包括: 从 URL 访问请求中携带的域名信息中提取出的特征; 和/或从 URL 访问请求中携带的 URL 参数中提取出的特征。

6. 根据权利要求5所述的方法, 所述特征包括以下特征中的多个的组合: 字符总数、字母总数、数字总数、符号总数、不同字符数、不同
15 字母数、不同数字数、不同符号数。

7. 一种 URL 攻击检测装置, 所述装置包括:

第一提取模块, 从 URL 访问请求中携带的信息中提取若干维度的特征;

- 计算模块, 将提取到的特征输入预设的 URL 攻击检测模型进行预测
20 计算, 得到所述 URL 访问请求的风险评分; 其中, 所述 URL 攻击检测模型为基于 Isolation Forest 机器学习算法训练得到的机器学习模型;

确定模块, 基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

8. 根据权利要求7所述的装置, 所述装置还包括:

- 25 第二提取模块, 从若干 URL 访问请求样本携带的信息中分别提取若干维度的特征; 其中, 所述若干 URL 访问请求样本均未被标记样本标签;

构建模块, 基于提取到的特征构建若干训练样本;

训练模块，基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得到所述 URL 攻击检测模型。

9. 根据权利要求 8 所述的装置，所述 URL 攻击检测模型包括基于 Isolation Forest 机器学习算法训练得到的 M 棵随机二叉树；

5 所述训练模块：

基于从所述若干训练样本中均匀抽样出的训练样本构建出 M 个训练样本子集；

10 从所述若干维度的特征中为各训练样本子集随机选择一分类特征作为根节点，以及在所述分类特征的最大取值和最小取值构成的取值区间中，为各训练样本子集随机选取一分类临界值；

将各训练样本子集中所述分类特征的取值大于所述分类临界值的训练样本，和所述分类特征的取值小于所述分类临界值的训练样本，分别分类为所述根节点的叶节点；以及，

15 将各叶节点中的训练样本作为新的训练样本子集，迭代执行以上分类过程，直到得到的各叶节点中的训练样本不可再分类时停止。

10. 根据权利要求 9 所述的装置，所述计算模块：

基于提取到的特征构建预测样本；

基于所述预测样本中各特征的取值，从根节点开始遍历各棵随机二叉树查找与所述预测样本对应的叶节点；

20 计算查找到的叶节点在各棵随机二叉树中的路径深度的平均值，并对所述平均值进行归一化处理，得到所述 URL 访问请求的风险评分。

11. 根据权利要求 7 所述的装置，所述信息包括：域名信息，和/或 URL 参数；所述若干维度的特征包括：从 URL 访问请求中携带的域名信息中提取出的特征；和/或从 URL 访问请求中携带的 URL 参数中提取出的特征。

12. 根据权利要求 11 所述的装置，所述特征包括以下特征中的多个的组合：字符总数、字母总数、数字总数、符号总数、不同字符数、不

同字母数、不同数字数、不同符号数。

13. 一种电子设备，包括：

处理器；

用于存储机器可执行指令的存储器；

5 其中，通过读取并执行所述存储器存储的与 URL 攻击检测的控制逻辑对应的机器可执行指令，所述处理器被促使：

从 URL 访问请求中携带的信息中提取若干维度的特征；

将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于

10 Isolation Forest 机器学习算法训练得到的机器学习模型；

基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。