文章编号:1671-4598(2023)10-0090-07 DOI:10.16526/j.cnki.11—4762/tp.2023.10.014 中图分类号:TP393 文献标识码:A

基于GBDT优化算法的局域网入侵定位与检测研究

蔡 娟[1],兰娅勋[1],刘 源[2]
(1.广州科技职业技术大学 信息工程学院,广州 510005;
2.南京理工大学 电子工程与光电技术学院,南京 210094)

摘要:入侵检测方法可以在局域网攻击造成广泛破坏之前发现攻击,并据此制定相应的防御措施;为保证局域网的运行安全,提出基于GBDT优化算法的局域网入侵定位与检测方法;考虑局域网的组成结构与工作原理,构建局域网数学模型;在该模型下,根据不同入侵类型的攻击原理,设置入侵检测标准;局域网实时运行数据采集与预处理,从时域和频域两个方面提取局域网的运行特征;利用GBDT优化算法构建局域网入侵分类器,匹配局域网运行数据特征,追踪局域网入侵源位置,最终得出入侵源定位与入侵状态、类型的检测结果;通过性能测试实验,发现与传统方法相比,优化设计方法的入侵定位误差降低了5.75 m,入侵类型与入侵数量的正确检测率分别提高13.8%和15.4%,即优化设计方法在定位与检测性能方面具有明显优势。

关键词:GBDT优化算法;局域网;入侵定位;入侵检测

Research on LAN Intrusion Location and Detection Based on GBDT Optimization Algorithm

CAI Juan[1], LAN Yaxun[1], LIU Yuan[2]
(1. College of Information Engineering, Guangzhou Vocational and Technical University of Science and Technology, Guangzhou 510005, China;
2. School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: Intrusion detection methods can detect local area network attacks before causing widespread damage, and develop corresponding defense measures accordingly. To ensure the operational security of the local area network, a local area network intrusion localization and detection method based on the GBDT optimization algorithm is proposed. Consider the composition structure and working principle of the local area network, and construct a mathematical model of the local area network. Under this model, intrusion detection standards are set based on the attack principles of different intrusion types. Real time operation data collection and preprocessing of the local area network, extracting the operational characteristics of the local area network from both time-domain and frequency-domain aspects. Using the GBDT optimization algorithm to construct a local area network intrusion classifier, matching the characteristics of local area network operation data, tracking the location of local area network intrusion sources, and ultimately obtaining the detection results of intrusion source localization, intrusion status, and type. Through performance testing experiments, it was found that compared with traditional methods, the optimized design method reduced the intrusion localization error by 5.75m, and improved the correct detection rates of intrusion types and intrusion quantities by 13.8% and 15.4%, respectively. This indicates that the optimized design method has significant advantages in localization and detection performance.

Keywords: GBDT optimization algorithm; LAN; intrusion location; intrusion detection

0 引言

局域网是由局部地区组成的一个区域网络,它的特点是它的分布面积是有限的,覆盖范围通常为方圆数公里,具有安装方便、成本节约、扩展方便等优点,使得它在各类办公室中得到了广泛的应用。在实际应用中,对局域网网络的安全性进行维护,可以对数据的安全性

收稿日期:2023-04-10; **修回日期**:2023-05-11。
基金项目:2022教育部协同育人项目(220601418293528)。
作者简介:蔡 娟(1983-),女,硕士,副教授。
兰娅勋(1981-),女,硕士,副教授。
引用格式:蔡 娟,兰娅勋,刘 源.基于GBDT优化算法的局域网入侵定位与检测研究[J].计算机测量与控制,2023,31(10):90-96.

进行有效的保障，确保局域网网络的正常、稳定运行。因为局域网具有移动性、传输性等特点，局域网极易遭受来自外部或网络自身的攻击，然而，在局域网中，常规的防火墙发挥的作用十分有限，无法对局域网提供良好的防护，从而使局域网面临潜在的威胁[1]。为了最大程度地保证局域网的运行安全，降低甚至消除非法入侵行为对局域网的影响，提出入侵定位与检测方法。入侵是非授权访问信息系统以及未经允许对信息系统进行操作，是对防火墙的合理补充，它可以帮助系统对付网络攻击。而入侵定位就是确定入侵攻击源的位置，为局域网的攻击防御提供辅助参考。

从目前入侵定位与检测方法的研究情况来看，文献[1]提出的基于天牛群优化与改进正则化极限学习机的网络入侵检测方法、文献[2]提出的基于机器学习的入侵检测技术和文献[3]中提出的基于超参数自动寻优的入侵检测方法发展较为成熟，其中文献[1]提出的入侵检测方法采用 LU 分解方法解决 RELM（正则化极限学习机）的权重矩阵问题，并采用 BSO（天牛群优化）方法实现权重与门限的联合优化；针对 BSO 方法易陷入局部极小化问题，拟采用 TentMapping 逆向学习、莱维飞行种群学习、动态突变等方法，以提高 BSO 方法的寻优能力。文献[2]提出的入侵检测方法应用了机器学习算法，构建支持向量机节点定位模型。确定当前阶段的网络通讯状况，建立了节点重要度的概念，并定义了具有较高重要度的节点为易受攻击目标，估计可能因不正常行为而导致的损失。而文献[3]中提出的入侵检测方法主要采用过采样技术解决原始数据中样本不平衡的问题．利用贝叶斯优化方法，对层叠的长－短时内存网络进行最优超参数值集的求解，实现对入侵检测的有效识别。将上述传统入侵检测方法应用到局域网环境中，发现输出结果存在明显的入侵定位误差与检测错误现象，为此引入 GBDT（gradient boosting decision tree，梯度提升决策树）优化算法。

GBDT 优化算法又称梯度提升决策树优化算法，它是一种将多个决策树组合在一起，然后将每一个决策树的结果相加得到最后的结果。GBDT 最优算法的基本思想是使得它在寻找多个可区分的特征及其组合方面有着天然的优势。利用 GBDT 优化算法，以局域网作为研究对象，优化设计入侵定位与检测方法，以期能够提升对局域网入侵行为的检测精度，间接的提高局域网的运行安全。

1　局域网入侵定位与检测方法设计

入侵检测是用来发现外部攻击与合法用户滥用特权的一种方法。局域网入侵定位与检测方法的运行原理如图 1 所示。

在图 1 所示的原理的支持下，局域网入侵检测的工作流程大体可以分为 3 个步骤，分别为数据收集、数据分析以及结果处理，通过当前局域网运行数据特征与不同入侵攻击下数据运行的标准特征进行匹配，确定当前局域网是否存在入侵行为，识别入侵类型以及入侵点位置。

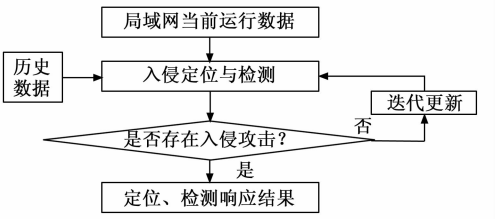

图 1　局域网入侵定位与检测原理图

1.1　建立局域网数学模型

从组成结构方面来看，局域网由网卡、网络终端、接入节点等部分组成。局域网大多采用聚类分层连接结构，网络由几个簇组成，每一个簇都包含了一个簇首和多个簇成员，簇头是由节点利用分簇算法自动选举产生的[2]。假设局域网中存在 n_p 个节点，任意两个节点之间存在连接链路，那么局域网中的链路数量可以表示为：

$$n_l = \frac{n_p(n_p-1)}{2} \tag{1}$$

构建的局域网数学模型设置 IEEE802.11g 协议作为数据传输协议，采用了正交频分多路复用 OFDM 编码技术，以克服传统 DSSS 编码方案中存在的多径衰减问题。不仅能满足 2.4 GHz 频段的数据传输速率，而且能保证与同样数量的机器设备相兼容。将传输协议融合到局域网结构模型中，得出局域网数学模型的构建结果。

1.2　设置局域网入侵检测标准

在构建的局域网数学模型下，通过模拟多种典型的网络入侵行为，确定不同入侵类型的作用原理，并将不同入侵行为下局域网实时数据的变化特征作为入侵检测的判定标准[3]。根据入侵攻击行为的后果，可以将入侵类型分为：拒绝服务攻击、中间人攻击、重放攻击等，其中局域网的拒绝服务攻击原理如图 2 所示。

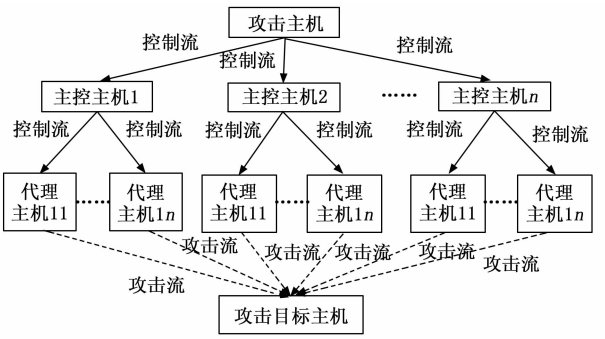

图 2　局域网拒绝服务攻击原理图

在拒绝服务攻击中，攻击者首先选取多台存在安全缺陷的主机作为主控主机，然后对其进行访问。在确定代理主机时，采用了与确定主主机相似的方式，不同之处在于，攻击者可以通过主机来间接地进行入侵。非法用户为了加强对局域网的攻击强度，会尽量增加局域网中傀儡主机和代理主机的数量，并在入侵攻击时向目标主机发送大量的恶意攻击程序与指令，促使代理主机执行实际的 DDoS 攻击

任务[4]。在 DDoS 攻击程序下，傀儡主机包括主控主机和代理主机两个部分，并由此形成攻击对象的傀儡网络，傀儡网络一般都处于被攻击方或被攻击方之外。攻击者在选定了主服务器和代理服务器之后，就可以通过对这些服务器的通讯进行攻击。另外，局域网的中间人攻击是指在通信双方的中间有一个侦听者，通信双方一直都以为他们是在直接通信，但实际上，他们所有交换的数据都是由这个侦听者转发的，即他们发出的消息，先发送到中间的侦听者，侦听者修改消息后再转发，一般情况下，攻击者会使用该方法获取用户名和密码等关键信息，再利用这些信息连接到合法 AP[5]。同理可以得出其他入侵攻击的模拟结果，并在入侵攻击程序下，实时采集局域网的实时运行数据，提取任意入侵攻击下的运行特征，将 i 类入侵攻击下的数据运行标准特征标记为 $\lambda_{\text{standard}}(i)$，以此作为局域网入侵攻击行为的检测标准。

1.3 局域网实时运行数据采集与预处理

采用 BPF 技术截获局域网的实时运行数据包，BPF 技术由网络接口和数据包过滤器两部分组成，网络接口的实现是由网络驱动来完成对网卡的传输，然后将传输到待处理的系统中。然后，分组过滤器根据用户自定义的对应规则，将不需要的分组剔除，以避免分组进入用户空间。利用包过滤来缓存有效的包，以供用户使用。BPF 的主要作用是实现对局域网传输数据包的拦截和筛选，然后将经过筛选的报文缓存到系统内核中，供用户从所提供的应用程序接口进行调用[6]。局域网传输数据包的具体截获过程如图 3 所示。

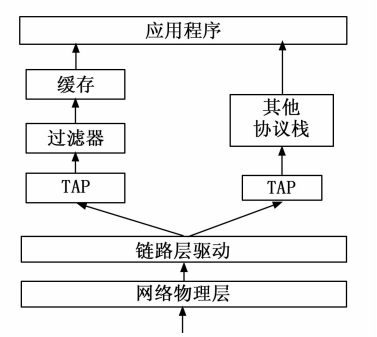

图 3 局域网传输数据包截获流程图

在入侵定位与检测过程中，需要对局域网中的数据变化情况进行分析，因此需要足够数量的数据作为支持，最终得出的局域网数据截获结果可以表示为：

$$N_{\text{LAN}}(x) = f \cdot \Delta t_{\text{catch}} \tag{2}$$

其中：f 为局域网数据的截获频率，Δt_{catch} 表示局域网运行数据的采集时间，公式（2）的输出结果 $N_{\text{LAN}}(x)$ 即为以 x 为传输内容的截获数据量[7]。为保证优化设计方法的入侵定位与检测精度，需要对初始截获的局域网数据进行预处理，预处理内容具体包括：冗余数据过滤、缺失数据补偿以及数据归一化，在冗余数据过滤过程中，首先计算初始捕获的任意两个数据之间的冗余度，计算公式如下：

$$r(x_i, x_j) = \frac{x_i \cdot x_j}{\| x_i \| \cdot \| x_j \|} \tag{3}$$

式（3）中，变量 x_i 和 x_j 分别为初始捕获数据中的第 i 和 j 个数据。若式（3）中的计算结果 $r(x_i, x_j)$，高于设置阈值 r_0，则说明 x_i 与 x_j 之间为冗余数据，需要删除其中任意一个数据，否则认为两者不为冗余数据，不需要执行过滤处理[8]。初始截获局域网传输数据中缺失部分的补偿过程可以量化描述为：

$$x'_k = \frac{x_{k-1} + x_{k+1}}{2} \tag{4}$$

其中：x_{k-1} 和 x_{k+1} 分别为缺失数据 x_k 的前后相邻数据，x'_k 为缺失数据的补偿结果。另外局域网捕获数据的归一化处理结果为：

$$x_{\text{normalization}} = \frac{x_{\max} - x}{x_{\max} - x_{\min}} \tag{5}$$

式中，$x_{\max}$ 和 $x_{\min}$ 为初始捕获局域网运行数据中的最大值和最小值[9]。将所有的局域网传输数据捕获结果依次代入到式（3）～（5）中，完成局域网实时运行数据的预处理，将最终的预处理结果幅值给初始截获数据。

1.4 提取局域网运行数据特征

以局域网传输数据的截获与预处理结果为处理对象，从时域和频域两个方面提取局域网的运行特征。设置局域网运行特征的提取向量为流量均值、流量峰值、流量波动因子等，其中流量均值的提取结果如下：

$$\lambda_{\text{mean value}} = \frac{\sum_{i=1}^{N_{\text{LAN}}} x_i^2}{N_{\text{LAN}}} \tag{6}$$

另外局域网流量峰值特征与流量波动因子特征的提取结果为：

$$\begin{cases} \lambda_{\max} = y_{\max}(x) \\ \lambda_{\text{undulate}} = \dfrac{\chi}{\sigma} \end{cases} \tag{7}$$

式中，变量 χ 和 σ 分别为方根幅值和绝对平均值，上述变量的求解公式为：

$$\begin{cases} \chi = \left(\dfrac{1}{N_{\text{LAN}}} \sum_{i=1}^{N_{\text{LAN}}} \sqrt{|x_i|} \right)^2 \\ \sigma = \dfrac{\sum_{i=1}^{N_{\text{LAN}}} |x_i|}{N_{\text{LAN}}} \end{cases} \tag{8}$$

将式（8）的计算结果代入到式（7）中，即可得出局域网运行流量峰值与波动因子特征的提取结果。同理可以得出其他所有局域网运行时域特征向量的提取结果。局域网运行数据的频域特征提取就是通过小波分解，提取局域网运行能量特征[10]。在实际的局域网频域特征提取过程中，首先利用离散小波变换技术对初始捕获数据进行 c 层分解，分解结果如下：

$$x_c(t) = \sum_{j=1}^{c} \sum \kappa_{\text{wavelet}} x_{\text{low}}(t) + \sum \kappa_{\text{approximate}} x_{\text{tall}}(t) \tag{9}$$

式中，κ_{wavelet} 和 $\kappa_{\text{approximate}}$ 分别为小波系数和近似系数，$x_{\text{low}}(t)$ 和 $x_{\text{tall}}(t)$ 为初始采集传输数据中的低频部分和高频部分，则局域网能量特征的提取结果为：

$$\lambda_E = \sum_{t=1}^{c} x_c(t) \cdot L(t) \tag{10}$$

其中：$L(t)$ 为 t 时刻局域网运行数据的信号长度。为方便局域网运行数据特征的统一处理与匹配，对提取的时域与频域特征作融合处理，最终得出的综合特征提取结果如下：

$$\lambda_{\text{draw}} = \bar{\omega}_1\lambda_E + \bar{\omega}_2\lambda_{\text{mean value}} + \bar{\omega}_3\lambda_{\max} + \bar{\omega}_4\lambda_{\text{undulate}} \tag{11}$$

式中，$\bar{\omega}_1$、$\bar{\omega}_2$、$\bar{\omega}_3$ 和 $\bar{\omega}_4$ 分别对应的是流量均值、流量峰值、流量波动因子以及能量特征向量对应的权重值[11]。由于局域网中的传输数据处于动态变化的状态，因此需要根据数据的采集频率对提取特征进行实时更新。

1.5　利用 GBDT 优化算法构建局域网入侵分类器

局域网入侵分类器构建的目的是为局域网入侵状态以及入侵类型的检测提供运行环境，为保证局域网入侵分类器输出的入侵类型检测结果，利用 GBDT 优化算法对分类器进行优化构建[12]。GBDT 优化算法是一种综合模式，其预测方法是将各个子树的预测值相加。GBDT 算法逐步产生一个决策子树来产生整片森林，并根据样本标签和现有林分预报结果的残差来构造新的子树。GBDT 优化算法的运行原理如图 4 所示。

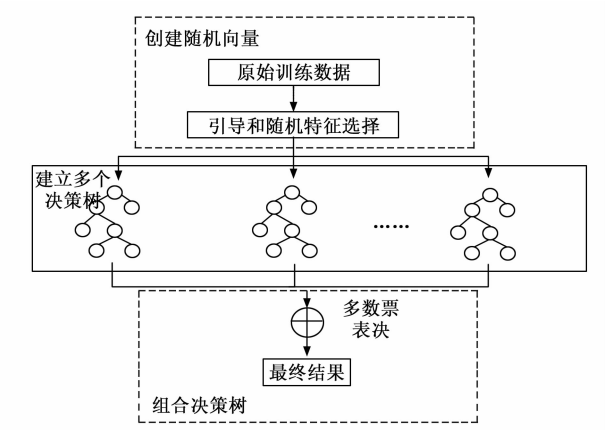

图 4　GBDT 优化算法原理图

GBDT 优化算法的运行大体可以分为决策树生成、负梯度拟合等步骤，如图 4 所示。定义提取的局域网运行数据特征作为 GBDT 优化算法的训练样本，在训练样本所在的输入空间中，对训练样本进行递归式分割，确定各子域的输出值，从而构造出一棵二叉决策树[13]。在决策树生成过程中，首先选择最优切分变量和切分点，分别标记为 b 和 q，用选定的 (b,q) 划分区域并决定相应的输出值：

$$\begin{cases} \beta_1(b,q) = \lambda_{\text{draw}} \mid \lambda_{\text{draw}}^{(b)} \leqslant q \\ \beta_2(b,q) = \lambda_{\text{draw}} \mid \lambda_{\text{draw}}^{(b)} > q \\ \xi_i = \dfrac{1}{N_{\beta\lambda_{\text{draw}} \in \beta(b,q)}} y_{\text{select}}(i) \cdot \beta_1(b,q) \cdot \beta_2(b,q) \end{cases} \tag{12}$$

式中，y_{select} 为最优切分变量和切分点选择的最优求解函数，N_β 代表区域划分数量，ξ 为区域划分的输出值，重复切分点选择、空间划分步骤，直至满足停止条件，将输入到 GBDT 优化算法中的局域网运行数据特征空间的划分成多个区域，并由此得出决策树的构建结果，可以量化表示为：

$$T = \sum_{i=1}^{n_\beta} \xi_i \cdot \lambda_{\text{draw}} \tag{13}$$

按照上述流程得出决策树的构建结果。采用加法模型和正向逐步算法来实现学习的优化过程。GBDT 优化算法通过对逻辑回归中的对数损失函数来实现分类工作[14]。为解决 GBDT 优化算法学习过程中存在的损失函数拟合的问题，用损失函数的负梯度拟合出一个回归树。如果损失函数是二次损失，那么这个负梯度就是目前分类器的残差值，因此，在损失函数是其他函数时，可以使用这负梯度来近似目前分类器的残差值[15]。第 k 轮的第 i 个样本的损失函数的负梯度表示为：

$$\delta = -\left[\frac{\partial T(y_i, y_{\text{study}}(\lambda_{\text{draw}}))}{\partial T(\lambda_{\text{draw}})}\right] \tag{14}$$

式中，y_{study} () 为任意一轮决策树的学习函数。通过损失函数的负梯度来拟合，解决分类器运行过程中的分类回归问题[16]。根据上述 GBDT 优化算法得出局域网入侵分类器的构建结果，构建分类器的工作逻辑如图 5 所示。

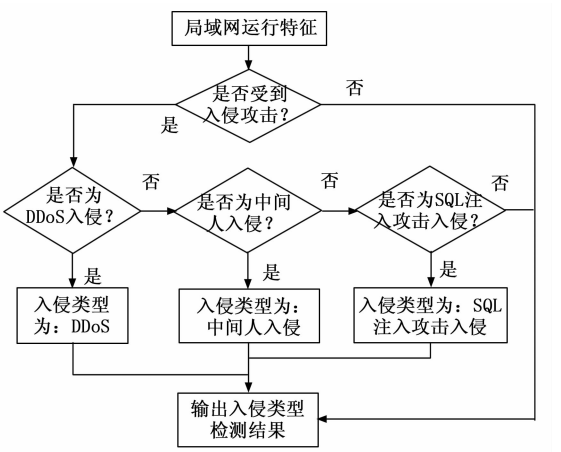

图 5　局域网入侵分类器工作逻辑图

局域网入侵分类器的数学表达式为：

$$F = \text{argmin}\beta(b,q) + \delta \tag{15}$$

将 GBDT 优化算法的学习结果以及相关运行参数代入到公式 (15) 中，即可得出局域网入侵分类器的最终构建结果。

1.6　匹配局域网运行数据特征

在构建局域网入侵分类器的支持下，通过提取局域网运行数据特征与设置入侵特征之间的匹配，判断当前局域网是否存在入侵攻击行为，并确定当前局域网的入侵攻击类型。局域网运行状态的特征匹配结果如下：

$$\gamma = F(\sqrt{\lambda_{\text{draw}} - \lambda_{\text{standard}}(i)})^2 \tag{16}$$

将 i 类入侵攻击特征与当前局域网的提取特征代入到公式（16）中，即可得出当前局域网与 i 类入侵攻击的匹配系数求解结果[17]。若计算得出匹配系数 γ，高于阈值 γ_0，说明当前局域网存在入侵攻击，且入侵类型为 i，否则进行下一入侵攻击类型的匹配，直到满足特征匹配条件为止，若当前局域网运行特征不与设置的任意一个入侵类型相匹配，则说明当前局域网不处于入侵攻击状态。

1.7 追踪局域网入侵源位置

针对处于入侵攻击状态下的局域网，需要确定入侵源的具体位置，并根据入侵源的移动情况进行追踪。采用距离向量技术执行局域网入侵源定位操作，即采用平均每跳距离来估算实际距离。首先将一个包含跳跃数段的数据包发送给邻接节点，并将该数据包发送给该邻接节点。一个被攻击的目标节点，其到每一个被攻击的节点都有一个最小的跳，并且被忽略掉了[18]。这个跳跃值被添加到 1，并被发送到邻近的节点。利用这种方式，局域网中的全部结点都可以记录到攻击结点的最小距离。入侵攻击节点的平均跳距可以表示为：

$$l_{ij}=\frac{\sum_{j\neq i}^{N_{\text{panel point}}}\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}}{\sum_{j\neq i}^{N_{\text{panel point}}}l_{\min}} \tag{17}$$

式中，(x_i,y_i) 和 (x_j,y_j) 分别为局域网中任意节点 i 和 j 的坐标，$l_{\min}$ 为局域网节点 i 和 j 之间的最小跳数，$N_{\text{panel point}}$ 表示局域网中包含的节点数量。每个攻击点在得到平均跳跃距离之后，会直接或间接地向外广播含有该平均跳跃距离的信息，而被接收到的信息只会被保存在它接收到的最初跳跃距离中[19]。接着，该未知节点将接收到的平均跳距之和乘以与之相连的任何一个攻击节点的最小跳数，从而获得该未知节点与之相连的任何一个攻击节点的估算距离，则未知入侵攻击节点与已知节点之间的距离可以表示为：

$$d(i,g)=l_{ig}\times n_{\text{pace}}+l_0 \tag{18}$$

其中：l_0 为首次接收到的平均跳距，g 为未知攻击节点，n_{pace} 为节点 i 与攻击节点之间需要的步数。那么根据入侵攻击节点与已知局域网节点之间的攻击关系，可以得出入侵节点的攻击结果为：

$$\begin{cases}x_g=x_i\cdot d_x(i,g)\\ y_g=y_i\cdot d_y(i,g)\end{cases} \tag{19}$$

式中，$d_x(i,g)$ 和 $d_y(i,g)$ 分别为节点间距离在水平和竖直方向上的分量，通过公式（19）的求解，即可得出当前局域网入侵源位置的定位结果。结合局域网实时运行数据的采集频率，重复上述入侵源定位流程，完成局域网入侵源的追踪与更新。

1.8 实现局域网入侵定位与检测

由于不同的入侵攻击类型的作用原理不同，因此产生的局域网运行特征存在明显区别，部分入侵可能不会对局域网的运行流量产生影响，而直接作用在网络传输协议上，因此除上述过程外，还需要对局域网的运行协议进行检测，主要就是对局域网传输数据的结构进行检测，最终将局域网入侵位置、类型的检测结果以可视化的形式输出，输出结果如下：

$$W=(Z(\gamma),(x_g,y_g)) \tag{20}$$

式中，$Z(\gamma)$ 为经特征匹配得出的局域网入侵状态检测结果[20]。按照上述流程，完成局域网入侵定位与检测工作。

2 性能测试实验分析

为了验证优化设计基于 GBDT 优化算法的局域网入侵定位与检测方法的定位与检测性能，以白盒测试为测试方式，设计性能测试实验。此次性能测试实验的基本思路为：在局域网测试环境下，设置入侵攻击点，以多个不同类型、不同强度的执行入侵攻击任务，利用优化设计的局域网入侵定位与检测方法输出检测结果，与设置的入侵攻击任务参数进行比对，验证优化设计方法输出结果与实际入侵情况之间的差距。

2.1 布设局域网测试环境

此次实验选择的局域网由 250 台主机和 3 台路由器组成。其中 1 台用来作为分析主机，能够执行优化设计的基于 GBDT 优化算法的局域网入侵定位与检测方法，另外随机选择局域网中的任意一台主机作为攻击主机，执行编写的攻击程序，对局域网中的其他节点主机进行入侵攻击。为了避免误差，无论是合法主机还是攻击主机，都使用了同样的硬件配置，使用的都是 WindowsX 的 SP2 操作系统。同时，两者均采用了 Realtek10/100 M 的适配网络卡，并采用了 TP-链接 10/100 M 的快速适配开关。

2.2 编写局域网入侵攻击程序

利用 MDK3 工具模拟局域网入侵攻击程序，该工具支持 DDOS 攻击、洪水攻击、中间人攻击等入侵程序。在 MDK3 工具支持下，DDoS 攻击程序的运行情况如图 6 所示。

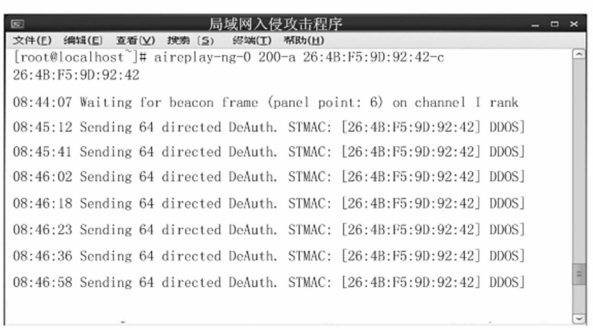

图 6 局域网入侵攻击程序运行图

通过无线网卡发射随机伪造的攻击信号，并根据需要设定攻击信号的工作频道，执行相应的攻击指令。为保证实验结果的可信度，此次性能测试实验设置多个入侵攻击任务，部分任务的设置情况如表 1 所示。

表 1　局域网入侵攻击任务表

任务编号	入侵类型	攻击强度等级	入侵节点数量	攻击目标节点
1	DDOS 攻击	Ⅰ级	10	节点 6
2	洪水攻击	Ⅰ级	45	节点 1
3	中间人攻击	Ⅰ级	27	节点 4
4	SQL 注入攻击	Ⅰ级	33	节点 12
5	DDOS 攻击	Ⅱ级	35	节点 6
6	洪水攻击	Ⅱ级	29	节点 1
7	中间人攻击	Ⅱ级	41	节点 4
8	SQL 注入攻击	Ⅱ级	18	节点 12

除表 1 中的入侵攻击任务信息外，还需要确定入侵节点的位置，实验共设置 200 个入侵攻击任务，其中攻击强度等级为Ⅰ级和Ⅱ级的任务均为 100 个。在局域网入侵攻击程序编写与运行过程中，确定局域网的入侵目标节点。另外，在入侵攻击程序编写过程中，需要添加强制控制指令，保证入侵攻击程序的可控性。

2.3　准备局域网运行数据样本

在不启动入侵攻击程序的情况下，将局域网调整至运行状态，执行多个数据通信与传输任务，通过 iptables 获取局域网的实时运行数据包，作为局域网入侵检测的参考数据。在此基础上，启动编写的入侵攻击程序，重复执行局域网的通信传输任务，按照相同的数据采集方式获取运行数据，作为入侵检测的测试数据。

2.4　实验结果与分析

利用编程工具对优化设计的基于 GBDT 优化算法的局域网入侵定位与检测方法进行开发，并将准备的局域网运行数据样本输入到优化设计方法对应的运行程序中，通过决策树的构建与学习、特征提取与匹配等步骤，得出当前局域网入侵定位与检测结果。图 7 表示的是局域网执行 1 号入侵攻击任务时输出的入侵定位与检测结果。

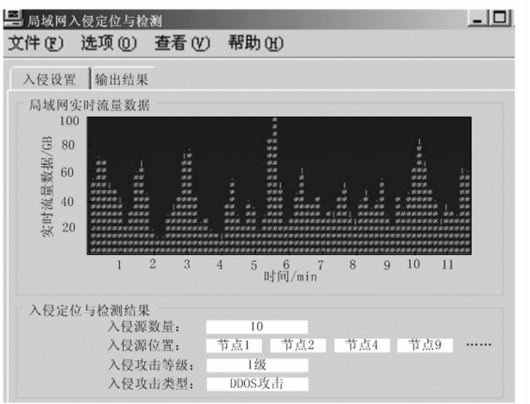

图 7　局域网入侵定位与检测结果

按照上述方式可以得出所有入侵攻击任务下的定位与检测结果，将优化设计方法的输出结果与设置攻击任务信息进行比对，判断优化设计方法输出结果是否准确。为了体现出优化设计方法在定位与检测性能方面的优势，设置传统的基于天牛群优化与改进正则化极限学习机的网络入侵检测方法和基于机器学习的无线传感网络通信异常入侵检测方法作为实验的对比方法，按照上述流程完成对比方法的开发与运行，并得出相应的输出结果。

首先，根据实验目的分别从入侵定位和入侵检测两个方面设置性能量化测试指标，其中入侵定位性能的测试指标为入侵定位误差，其数值结果为：

$$\varepsilon_{seat} = |x_g - x_{set}| + |y_g - y_{set}| \tag{21}$$

式中，(x_{set}, y_{set}) 为设置入侵攻击源的实际位置，根据局域网的规模，局域网入侵位置以米为计量单位。统计优化设计方法输出的入侵源定位结果，通过与设置入侵原位置的对比，得出反映入侵定位性能的测试结果。表 2 表示的是 1 号入侵攻击任务下入侵节点定位性能的测试结果。

表 2　局域网入侵定位性能测试数据表

节点编号	设置入侵节点位置		RELM-BSO 输出节点位置		基于机器学习输出节点位置		基于 GBDT 入侵节点位置	
	x	y	x	y	x	y	x	y
1	196	44	191	40	194	41	195	44
2	202	89	198	82	200	87	201	88
3	103	153	100	150	101	151	102	152
4	152	214	147	210	150	212	152	214
5	96	209	91	202	93	205	95	208
6	171	185	166	181	170	183	170	185
7	83	193	79	189	81	191	82	193
8	52	212	48	207	50	210	52	212
9	111	36	107	31	108	33	110	35
10	149	97	144	92	146	94	148	97

基于公式（21）计算节点位置偏差，统计结果如表 3 所示。

表 3　各算法的节点位置偏差值统计　　m

节点编号	RELM-BSO 输出节点位置	基于机器学习输出节点位置	基于 GBDT 入侵节点位置
1	9	5	1
2	11	4	2
3	6	6	2
4	9	7	0
5	12	7	2
6	9	3	1
7	8	4	1
8	9	4	0
9	9	6	2
10	10	6	1
均值	9.2	4.7	1.2

将表 2 中的数据代入到式（21）中，得出两种对比方法的平均入侵定位误差分别为 9.2 m 和 4.7 m，优化设计方法输出入侵定位误差的平均值为 1.2 m。

其次，设定入侵检测性能的测试指标分别为：入侵类型正确检测率和入侵数量正确检测率，上述指标的测试结果如下：

$$\begin{cases}\eta_{\text{type}}=\dfrac{n_{\text{type}}}{n_{\text{task}}}\times 100\%\\ \eta_{\text{quantity}}=\dfrac{n_{\text{Attack Node}}}{n_{\text{panel point}}}\times 100\%\end{cases}\tag{22}$$

式（22）中，变量 n_{type} 和 n_{task} 分别表示的是正确检测的入侵类型数量和设置的入侵任务总数量，$n_{\text{Attack Node}}$ 和 $n_{\text{panel point}}$ 对应的是成功识别出的入侵节点数量以及设置入侵节点的总数量。最终计算得出入侵定位误差越小、入侵类型正确检测率和入侵数量正确检测率越高，说明对应方法的定位与检测性能越优。通过公式（22）的计算，得出入侵类型正确检测率和入侵数量正确检测率的测试对比结果，如图 8 所示。

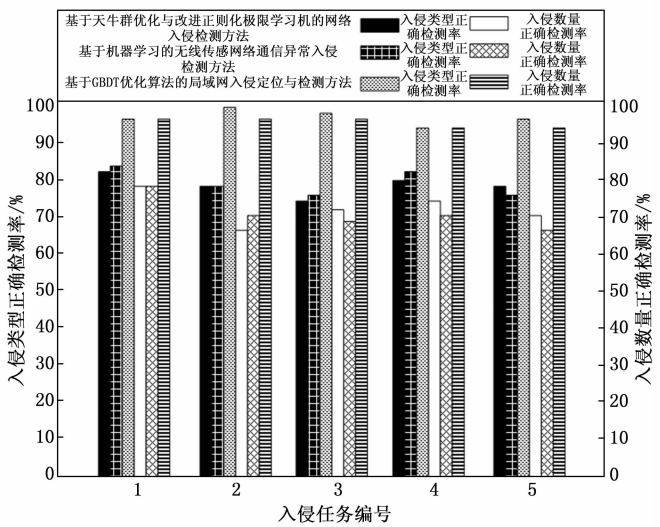

图 8 局域网入侵类型与数量检测性能测试对比结果

从图 8 中可以直观地看出，与传统入侵定位与检测方法相比，优化设计方法的入侵类型正确检测率和入侵数量正确检测率更高，入侵类型正确检测率提高约 13.8%，入侵数量正确检测率提高约 15.4%。

3 结束语

在非法攻击迅速发展的大背景下，攻击种类及其破坏程度急剧增加，为满足局域网在安全方面的需求，提出了局域网的主动防御技术，它的技术核心就是通过动态保护措施来保证系统的安全，从而能够及时地发现可局域网的异常运行情况，并针对入侵攻击进行控制和反击。该方法最大的优点是能够对新的攻击进行自适应地分析和学习。在主动防御技术中，入侵检测是一个非常重要的组成部分，它是通过从计算机网络或主机系统的关键点上收集信息，并对这些关键点信息展开分析，从而识别出局域网是否有存在企图入侵、正在进行入侵或已经入侵的行为。在未来，入侵检测方法将会向着高速实时化、智能化等方向发展。在此次研究中，通过 GBDT 优化算法的应用，实现局域网入侵定位与检测方法的优化。从实验结果中可以看出，优化设计方法具有更高的定位精度和检测精度，对于提高局域网的主动防御水平以及网络运行安全性能具有重要意义。

参考文献：

[1] 王振东，刘尧迪，杨书新，等．基于天牛群优化与改进正则化极限学习机的网络入侵检测［J］．自动化学报，2022，48（12）：3024－3041.

[2] 肖 衡，龙草芳．基于机器学习的无线传感网络通信异常入侵检测技术［J］．传感技术学报，2022，35（5）：692－697.

[3] 刘会鹏，周治平．基于超参数自动寻优的工控网络入侵检测［J］．信息与控制，2021，50（4）：427－434.

[4] 张永康，尚 盈，王 晨，等．分布式光纤入侵信号检测与识别［J］．光电工程，2021，48（3）：19－36.

[5] 刘奇旭，王君楠，尹 捷，等．对抗机器学习在网络入侵检测领域的应用［J］．通信学报，2021，42（11）：1－12.

[6] 马泽煊，李 进，路艳丽，等．融合 WaveNet 和 BiGRU 的网络入侵检测方法［J］．系统工程与电子技术，2022，44（8）：2652－2660.

[7] 杨 涛，叶西宁．网络入侵检测算法 SPCA-ERoF［J］．计算机工程与设计，2021，42（2）：356－362.

[8] 马静怡，崔昊杨，张明达，等．基于改进 Faster RCNN 的小尺度入侵目标识别及定位［J］．中国电力，2021，54（3）：38－44.

[9] 陈志华，黄志宏．基于知识图谱的激光通信网络入侵攻击源定位方法［J］．应用激光，2022，42（7）：118－124.

[10] 李 俊，夏松竹，兰海燕，等．基于 GRU-RNN 的网络入侵检测方法［J］．哈尔滨工程大学学报，2021，42（6）：879－884.

[11] 王艺霏，莫 爽，吴文睿，等．基于内外卷积网络的网络入侵检测［J］．北京邮电大学学报，2021，44（5）：94－100.

[12] 关 生，周延森．拆分降尺度卷积神经网络入侵检测方法［J］．科学技术与工程，2022，22（36）：16108－16115.

[13] 田桂丰，单志龙，廖祝华，等．基于 Faster R-CNN 深度学习的网络入侵检测模型［J］．南京理工大学学报，2021，45（1）：56－62.

[14] 陈 卓，吕 娜，陈 坤，等．基于时空图卷积网络的无人机网络入侵检测方法［J］．北京航空航天大学学报，2021，47（5）：1068－1076.

[15] 郭志民，周劼英，王 丹，等．基于 Transformer 神经网络模型的网络入侵检测方法［J］．重庆大学学报，2021，44（11）：81－88.

[16] 魏明军，张鑫楠，刘亚志，等．一种基于 SSA-BRF 的网络入侵检测方法［J］．河北大学学报（自然科学版），2022，42（5）：552－560.

[17] 汪祖民，王冬昊，梁 霞，等．基于 DBSCAN _ GAN _ XGBoost 的网络入侵检测方法［J］．郑州大学学报（工学版），2022，43（3）：44－51.

[18] 李 晶，黄 杰，朱国威，等．基于自适应一维 CNN 的网络入侵检测方法［J］．武汉大学学报（工学版），2022，55（11）：1176－1185.

[19] 贺佳星，王晓丹，宋亚飞，等．CWGAN-DNN：一种条件 Wasserstein 生成对抗网络入侵检测方法［J］．空军工程大学学报（自然科学版），2021，22（5）：67－74.

[20] 田桂丰，单志龙，廖祝华，等．基于空间降维和多核支持向量机的网络入侵检测［J］．济南大学学报（自然科学版），2021，35（4）：365－369，375.