南京郵電大學
Nanjing University of Posts and Telecommunications

用于入侵检测的长短期记忆递归神经网络分类器

1023040823　　何润杰

目录 CONTENT

南京郵電大學 Nanjing University of Posts and Telecommunications

01

背景

Part one

背景

由于信息和通信技术的进步，通过在线共享信息的情况有所增加，这带来了新的附加价值。因此，各种在线服务应运而生。然而，随着互联网连接点的增加，网络安全的威胁也在不断增加。其中，入侵检测系统（IDS）是当今重要的安全问题之一。

02

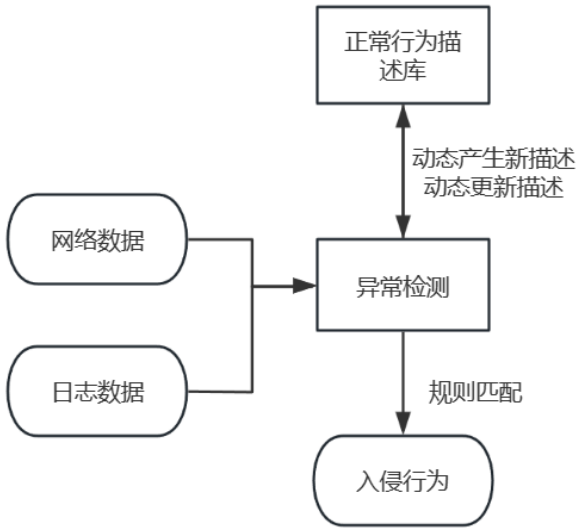
IDS检测原理及主要方法

Part two

基本原理

异常检测基本原理

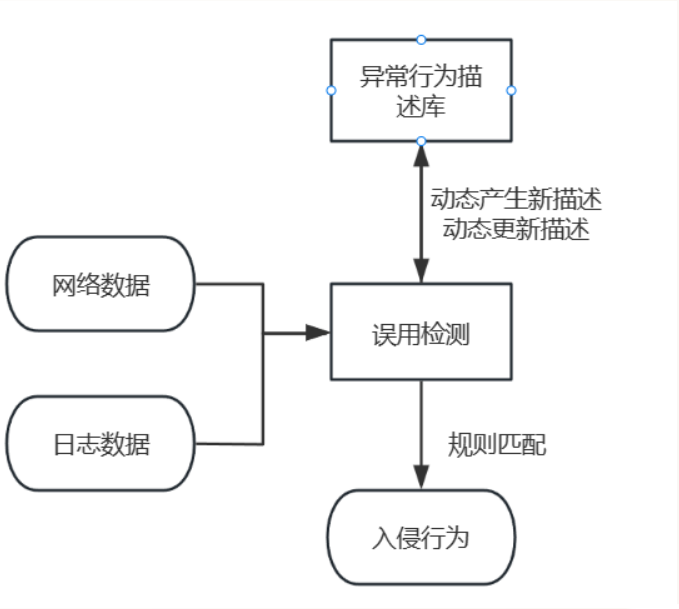
异常检测技术又称为基于行为的入侵检测技术，用来识别主机和网络中的异常行为。该技术假设攻击与正常合法的活动有明显的差异。这种方法可以发现未知攻击，但容易产生误报，需要较高的计算和存储资源。



基本原理

误用检测基本原理

误用检测技术又称为基于知识的入侵检测技术，通过匹配已知攻击特征或签名来检测攻击。这种方法依赖于已有的攻击知识库，能够快速准确地检测已知攻击，但对未知攻击和变种攻击无能为力。



主要方法

异常检测主要方法

基于时间序列分析的方法：

- **自回归积分滑动平均模型（ARIMA）：**用于时间序列数据，分析数据的时间依赖性，识别异常点。
- **长期短期记忆网络（LSTM）：**一种递归神经网络（RNN），特别适用于处理和预测时间序列中的长期依赖关系，用于检测行为序列中的异常。

基于统计的方法：

- **阈值检测：**设定一个或多个阈值，超过阈值的行为被视为异常。例如，设定CPU使用率的上限，如果超过该上限则触发警报。
- **统计模型：**建立正常行为的统计分布，如均值和标准差。行为偏离统计分布一定范围内的被视为异常。

主要方法

误用检测主要方法

基于签名的方法：

- **模式匹配：**使用预定义的攻击签名库，通过匹配网络流量或系统日志中的特定模式来检测攻击。例如，防病毒软件使用签名匹配检测已知病毒。
- **正则表达式匹配：**利用正则表达式定义复杂的攻击特征，匹配数据流中的相应模式。

基于协议分析的方法：

- **协议解码：**解析网络协议，检查是否存在违反协议规范的行为，如非法的报文格式或异常的协议字段值。可用于检测网络攻击，如TCP协议中的SYN洪水攻击。

03

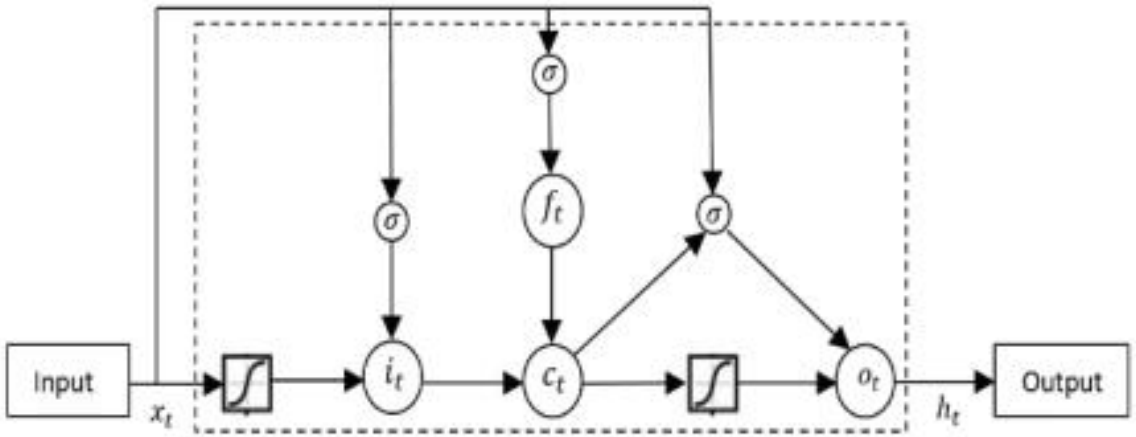
长短期记忆网络

Part three

异常检测为什么可以使用LSTM来实现

异常检测可以使用基于时间序列分析的方法，因为许多系统和数据在现实世界中都具有时间序列的特性，即它们随着时间的推移产生变化。基于时间序列分析的方法可以帮助识别数据中的趋势、周期性变化和异常事件。而LSTM是一种适用于序列数据的循环神经网络（RNN）变体，特别擅长处理时间序列数据。在异常检测中，许多数据都具有时间序列特性，如系统日志、网络流量、传感器数据等，LSTM可以有效地捕捉数据中的时间相关性和序列模式。

LSTM介绍



$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$$

$$c_t = f_t c_{t-1} + i_c tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)$$

$$h_t = o_t tanh(c_t)$$

输入门(it)：输入门控制着当前时间步的输入数据对细胞状态的影响程度。
遗忘门(ft)：遗忘门决定了上一个时间步的记忆状态对当前时间步的影响程度。
细胞状态更新(ct)：细胞状态 ct 在每个时间步都会被更新，通过遗忘门控制上一个时间步的记忆保留程度，通过输入门控制当前时间步的输入信息融合。
输出门(ot)：输出门决定了当前时间步的隐藏状态 ht对最终输出的影响程度。
隐藏状态(ht)：隐藏状态 ht 是 LSTM 的主要输出，包含了当前时间步的信息，通常用于后续的任务。

通过上述结构，LSTM在处理序列数据时可以有效地捕捉长期依赖关系，避免了传统RNN中的梯度消失和梯度爆炸问题，提高了模型的学习和泛化能力。

04

数据集

Part four

数据集

在许多研究中，KDD Cup 1999数据集被用来衡量IDS的性能。虽然数据集是旧的，但比较IDS模型是有益的。因为在相同的数据集上有很多性能测量结果。这也是我们选择KDD Cup 1999数据集的主要原因。
数据集中有4,898,431个网络流量，每个流量有41个特征。并根据其特点对22种攻击进行了分类。攻击类型有以下四种。DoS攻击会耗尽目标服务器的资源，使目标服务器无法提供任何服务。R2L攻击允许未经授权的远程访问。U2R攻击试图获取超级用户权限。Prob(探测攻击)用于发现目标服务器的漏洞。
KDD Cup 1999 数据集由于年代久远，有些特征和攻击类型可能不再符合当今网络环境的实际情况。因此，在使用这个数据集进行研究和评估时，需要结合实际情况进行分析和调整，以确保研究结果的有效性和可靠性。

05

实验和结论

Part five

实验结论

在使用训练数据集之前，将所有实例归一化到 0 到 1 之间。输入向量包含 41 个特征，输出向量由 4 种攻击类型和 1 种非攻击类型组成。因此，输入维度为 41，输出维度为 5。我们在隐藏层中应用 LSTM 架构。时间步长、批次大小和训练轮数分别为 100、50 和 500。最后在输出层使用 softmax，并使用随机梯度下降（SGD）作为优化器。损失函数为均方误差（MSE）。经过实验调整好超参数，并将结果与其他分类器算法进行比较。

Comparison with other algorithms

	DR(%)	FAR(%)	Accuracy(%)
GRNN	59.12	12.46	87.54
PNN	96.33	3.34	96.66
RBNN	69.83	6.95	93.05
KNN	45.74	46.49	90.74
SVM	87.65	6.12	90.4
Bayesian	77.6	17.57	88.46
LSTM-RNN	**98.88**	**10.04**	**96.93**

可以看出尽管误报率比其他算法略高，但检测率和准确率的百分比是最好的。