$\mathrm{Real}_T ime Multistep Attack Prediction Based on Hidden Markov Models$

海华

2020 年 7 月 12 日

1 总结

本论文提出了一种基于隐马尔可夫模型的方法来利用IDS警报预测多步攻击。IDS日志将告警处理为observes定义了tag和severity，其中tag是CVE报告中定义的漏洞描述提取词，severity为日志中告警等级字段。

框架设计包含部分：多步攻击文件（Multistep Attack files，pcap格式，用于HMM模型训练阶段），离线训练阶段，HMM配置文件（离线训练生成的参数，参数涵盖了状态转移可能性矩阵和马尔科夫状态链中状态的数目，主要用于预测阶段），预测模型利用HMM配置文件对IDS日志中的告警序列进行预测，计算攻击的可能性。

隐马尔可夫模型抽象成一个，第一个参数表示基于IDS告警和CVE报告的观察序列。第二个参数表示攻击状态。第三个参数表示状态转换可能性矩阵。第四个参数表示 an observation probability matrix，其实就是observation到S的可能性矩阵。第五个参数表示初始状态可能性向量。增加H向量表示每一步攻击状态的平均告警数。增加H后表示的六元祖向量可用于计算最后的攻击可能性。训练阶段分别采用监督学习方式和无监督学习方式。其中可以学习一下Baum-Welch算法是最常用的算法。还有其他的最大似然估计、梯度下降算法等。最后通过构建的六元祖结合训练的配置文件参数计算状态可能性和攻击可能性。最后论文的第7部分讲解了具体算法的实现实例，第八部分讲解了不同攻击场景构建的攻击配置文件。

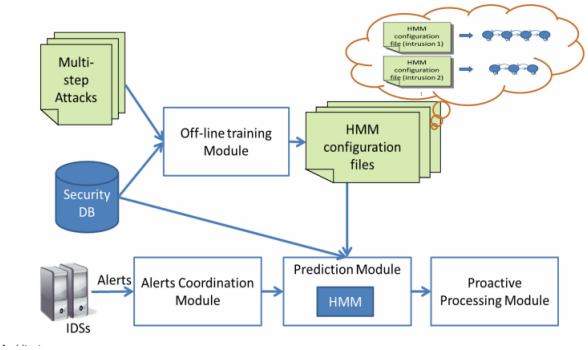

图 1: 框架

本文创新：

1 构建的六元祖模型，在原隐马尔可夫模型的基础上引入了状态的平均告警数目。

2 observation 状态基于CVE报告和日志告警等级。

3 本文的实验对性能、状态和攻击可能性、分布式攻击场景中算法的应用进行了分析。