SHADEWATCHER【反向推荐】

limitation:

- 误报率高；
- 依赖专家知识；
- 只能生成粗粒度的检测标志；

威胁检测任务和推荐系统之间的结构相似性；将系统实体之间的交互映射到user-item交互的概念之中；

- 预测系统实体对其所交互的其他实体的偏好，进而识别网络威胁；
- 利用GNN来建模审计日志的系统实体之间的高阶连接信息；
- 制定动态更新策略，从而更好的降低误报率；

1 introduction

溯源数据中的丰富的上下文信息使得安全人员能够对系统事件进行因果分析，从而进行入侵检测、依赖追踪、以及安全事件推理。

威胁检测/攻击检测是攻击调查（attack investigation）的第一步。

现有的基于溯源数据的攻击检测方法可以分为三类：

1. Statistics-based 基于统计方法：通过审计记录（事件）在溯源图中的罕见程度（出现频率极小）来量化该事件的嫌疑度（suspicious degree）；

   不足：会将很多出现频次很少但是仍然是正常的系统行为判定为异常事件，即产生误报；

   原因：仅考虑系统实体之间的有向连接，而没有考虑本身的语义或者事件的语义；

2. Specification-based 基于规则的方法：将审计日志记录与已有的攻击模式或者检测规则进行匹配；【Holmes、Rapsheet、SLEUTH、POIROT】

   不足：规则匹配的时间开销大；攻击模式规则的覆盖度很难保证全面；

3. Learning-based 基于学习的方法：通过机器学习技术建模正常行为，进而检测与正常行为存在偏差的异常行为；【Unicorn、ProvDetector、ATLAS】

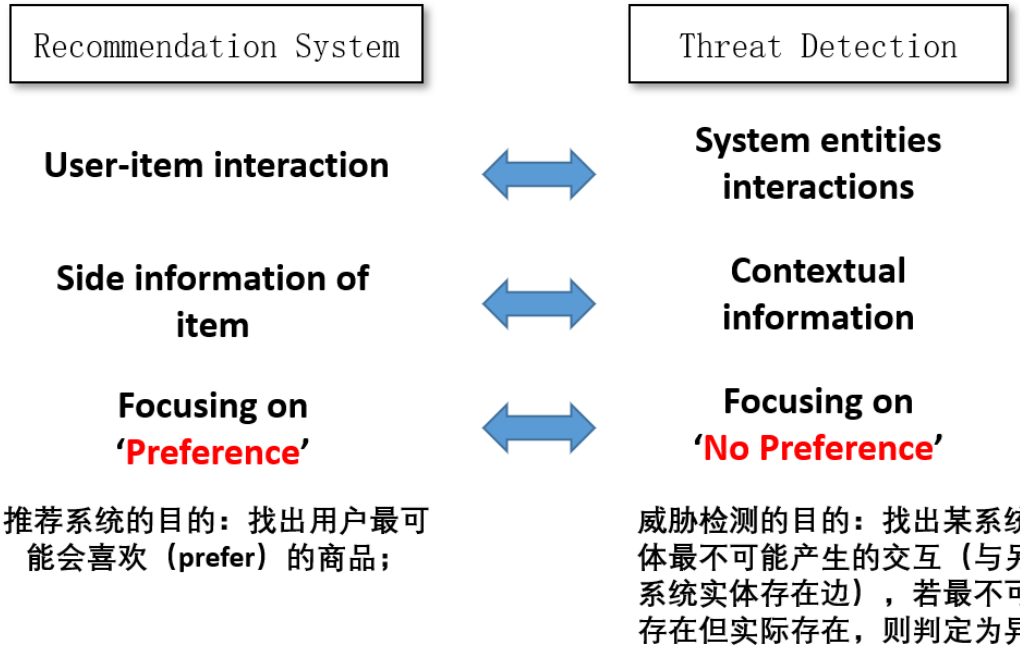
motivation

Threat actors typically induce unwanted behaviours with unintended system entity interactions deemed suspicious to analysts.

攻击者通常通过安全人员所认为的可疑的非预期的系统实体交互来进行一些非预期的（异常的）行为；【表明本文关注的重点在于：“***系统实体之间的非正常的交互***”，也就是异常的事件；】

- 系统中的预期的正常交互形成的行为规范和准则，组成了最可能被观察到的行为；[出现频率多]
- 系统中的非预期的系统实体之间的交互，这类交互行为与正常行为具有较大偏差；

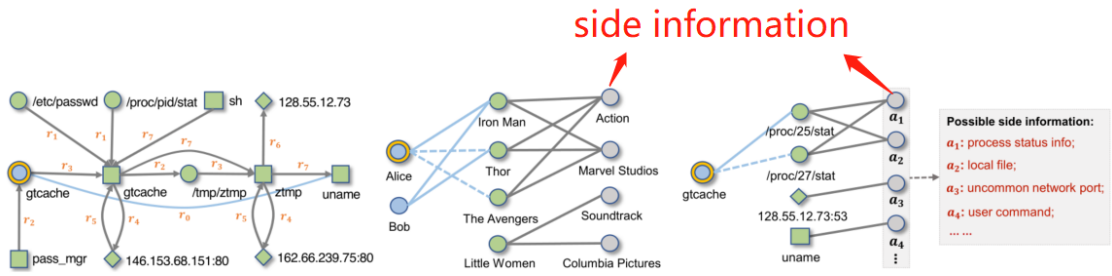
Cyber threats can be revealed by determining how likely a system entity would interact with another entity



SHADEWATCHER

- 利用上下文感知的嵌入模型，获取溯源图中实体的side information；【embedding模型】
- 利用基于GNN的推荐系统建模系统实体的交互行为【GNN的消息传播机制】，从而依据实体行为做出推荐；
- 人为反馈提供监督以实现动态更新；

2 Background and Motivation



(a) System Provenance Graph (b) Movie Recommendation (c) High-order Information in Audit Records

shadewatcher的关键就在于系统实体之间的交互。——观察到的***可能性***较低的交互行为就可以认为是潜在的网络威胁。

例如，上图a中，检测该后门攻击就转变为了——gtcache最不可能与哪个实体产生交互，若实际上产生了交互，则该交互行为是潜在的异常行为；

3 Problem definition

溯源图：

将系统事件转化为（src、relation、dst）的模式的溯源图；

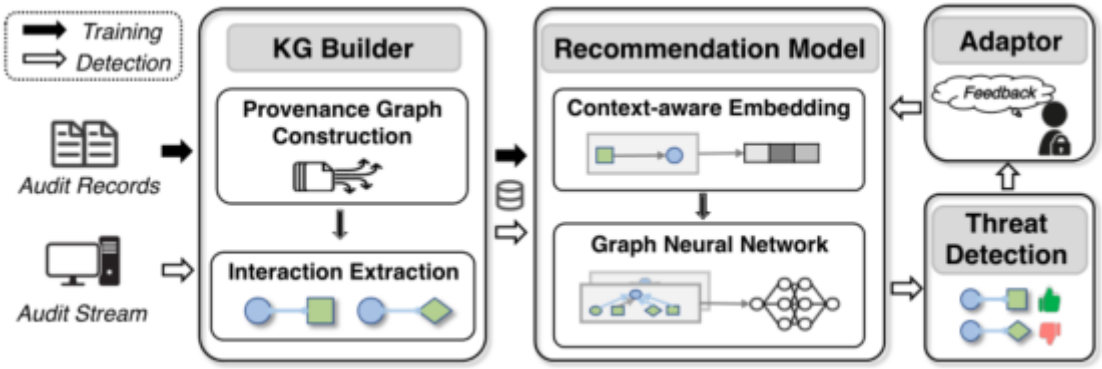
系统实体交互图

将系统实体之间的交互建模为二部图，不仅包括显式的数据依赖关系，还包括隐式的控制依赖关系；

知识图谱

将溯源图和系统实体交互图进行合并，构成最终的Knowledge Graph

4 System Overview



- KG Builder: 构建溯源图、二部图，合并为KG；
- Recommendation Model：使用正常数据训练基于GNN的推荐模型；
- Threat Detection：输入待检测交互数据，若预测概率大于设定阈值，则判定为异常；
- Adaptor：当系统行为发生明显改变后，通过人为干预对误报结果进行标定并反馈给系统，从而使得推荐系统能够实现一定的自适应；

provenance graph construction

事件去噪，保留攻击相关的信息；

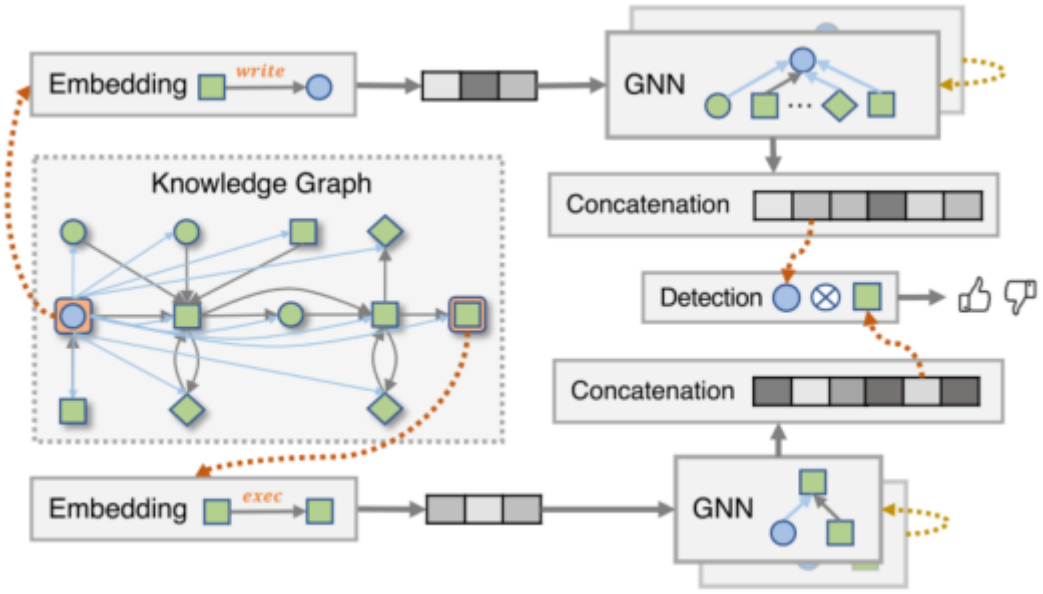
interaction extraction

跟watson一样的子图划分，找出溯源图中的数据对象节点（data objects），进而依据每个子图，可以提取出数据对象节点与系统实体节点之间的交互行为，将其构成二部图；

combining provenance graph and bipartite graph

PG由一系列三元组组成，BG也由一系列三元组组成，将这两部分的三元组进行merge，得到最终的KG

5 Recommendation Model



A.建模一阶信息

利用TransR算法对KG进行embedding，从而得到系统实体的语义和行为相似性的数值表示；

$$\mathcal{L}_{first} = \sum_{(h,r,t)\in\mathcal{G}_K} \sum_{(h',r',t')\notin\mathcal{G}_K} \sigma(f(h,r,t) - f(h',r',t') + \gamma),$$

B.建模高阶信息

TransR获得节点的embedding表示后，作为GAT的节点初始化embedding，进而进行GNN的多跳邻居节点的信息聚合；

C.威胁检测

1. 将系统实体h的GNN的不同层的计算结果进行拼接： $\mathbf{z}_h^* = \mathbf{z}^{(0)} || \cdots || \mathbf{z}_h^{(L)}.$
2. 给定一个交互（h, interact, t），计算h的embedding表示和t的embedding表示的内积：

   $\hat{y}_{ht} = \mathbf{z}_h^{*\top} \mathbf{z}_t^*.$ ，若乘积大于一个预定义的阈值，则将该交互标记为可能的网络威胁；

   $$\mathcal{L}_{higher} = \sum_{(h,r_0,t)\in\mathcal{G}_K} \sum_{(h',r_0,t')\notin\mathcal{G}_K} \sigma(\hat{y}_{ht} - \hat{y}_{h't'}),$$
3. higher损失函数中，通过对KG中已有的交互行为的头尾实体embedding的内积与不存在的交互行为的头尾实体的embededing的内积做减法。【正常的交互的内积尽可能小；异常的交互的内积尽可能大；】

D.模型适应

若系统产生误报，则将该误报的系统交互人为标注为良性交互后，重新作为训练数据输入推荐模型当中；

6 Implementation

learning_rate：0.001

embedding size：32

2-layer GNN with embedding size as 32 and 16

threshold：-0.5

7 Evaluation

DARPA TRACE数据集，实际上是所有的系统实体交互三元组，训练集、验证集和测试集比例为8：1：1；

- shadewatcher要求训练集为attack-free数据，即正常情况下的系统实体交互行为数据，因为shadewatcher本质上还是异常检测的思路；
- 验证集同样也不包含恶意交互；
- 论文作者对数据集中的攻击相关的事件所导致的系统实体之间的交互进行***手工标注***；

Shadewatcher的推荐模型的训练、验证和测试的过程中，全部使用正常的良性数据样本；

$$TN = \frac{\text{将良性交互预测为良性交互的数量}}{\text{良性交互的总数}}$$

$$FP = \frac{\text{将良性交互预测为异常交互的数量}}{\text{良性交互的总数}}$$

TN要尽可能大、FP要尽可能小

Appendix C——训练阶段的负采样

Appendix E——攻击场景

对应darpa e3groundtruth中的3.2、3.12、3.15、4.9