计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.191200172

基于因果知识和时空关联的云平台攻击场景重构

王文娟　杜学绘　任志宇　单棣斌
中国人民解放军战略支援部队信息工程大学　郑州 450001

摘　要　云计算环境下的攻击行为逐步表现出隐蔽性强、攻击路径复杂多步等特点,即一次完整的攻击需要通过执行多个不同的攻击步骤来实现最终目的。而现有的入侵检测系统往往不具有必要的关联能力,仅能检测单步攻击或攻击片段,难以发现和识别多步攻击模式,无法还原攻击者完整的攻击渗透过程。针对这一问题,提出了基于因果知识和时空关联的攻击场景重构技术。首先,利用贝叶斯网络对因果知识进行建模,从具有IP地址相关性的告警序列中发掘出具有因果关系的攻击模式,为后续关联分析提供模板依据。然后,借助因果知识网络,从因果、时间和空间多维度上对告警进行关联分析,以发现潜在的隐藏关系,重构出高层次的攻击场景,为构建可监管、可追责的云环境提供依据和参考。

关键词:云计算;攻击场景;告警关联;因果知识网络;时空关联
中图法分类号　TP309

Reconstruction of Cloud Platform Attack Scenario Based on Causal Knowledge and Temporal-Spatial Correlation

WANG Wen-juan,DU Xue-hui,REN Zhi-yu and SHAN Di-bin
PLA Strategic Support Force Information Engineering University,Zhengzhou 450001,China

Abstract　Attack behavior in cloud computing environment gradually shows characteristics of strong concealment and complex multi-step,that is,a complete attack needs to execute some different attack steps to achieve the final goal. However,the existing intrusion detection system usually does not have the necessary ability of correlation,and can only detect single-step attack or attack fragment,so it is difficult to find and identify multi-step attack,and unable to restore attackers' attack process completely. To solve this problem,this paper proposes an attack scenario reconstruction technique based on causal knowledge and space-time correlation. Firstly,the bayesian network is used to model the causal knowledge,and the causal attack patterns are extracted from the alerts with IP address correlation,so as to provide template basis for the subsequent correlation analysis. Then,on the basis of causal knowledge network,alert correlation is conducted from the perspectives of causal,temporal and spatial dimensions to discover potential hidden relationships,and high-level attack scenarios are reconstructed to provide basis and reference for building a cloud environment that can be monitored and accountable.

Keywords　Cloud computing,Attack scenario,Alert correlation,Causal knowledge network,Temporal-spatial correlation

1　引言

随着云计算[1]的不断发展和广泛应用,其安全问题日益剧增。云计算环境下的攻击行为(DDOS攻击、APT攻击等)[2]表现出隐蔽性好、复杂多步等特点,具体来说:1)一个完整的攻击过程是由多个具有不同目的的单步攻击组成,达到某个目的之后继续实施下一个攻击,攻击步骤之间存在因果依赖关系;2)攻击步骤间的因果关系存在着不确定性因素,无法确定某攻击是否一定会成功,前一攻击步骤成功实施能否导致下一攻击步骤的发生等[3];3)多步攻击具有序列性和时间性特征[4],各攻击步骤之间具有一定的规律和顺序,并且这些攻击步骤会在一定的时间间隔内完成;4)多步攻击是受拓扑空间约束的,攻击步骤之间在IP地址分布上总是具有关联性,如前一攻击步骤的目标IP地址就是下一攻击步骤的源IP地址。可见,多步攻击的步骤之间存在因果依赖关系、概率推理关系、时间序列关系以及空间关联关系。如何揭示出攻击步骤之间隐藏的逻辑关系并重构出攻击场景,是云计算安全面临的难题之一。攻击场景指呈现攻击者攻击渗透过程的"画面",通过"画面"能够获悉攻击者发动攻击的起点、各个攻击步骤之间的关系以及攻击者的攻击动机和攻击目标等信息,攻击场景常常用"图"的形式表示[5]。基于此,攻击场景重构就是将单步攻击或攻击片段按照一定的逻辑关系进行关

到稿日期:2019-12-30　返修日期:2020-04-23　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金项目(61802436);国家重点研发计划课题(2016YFB050190104)
This work was supported by the National Natural Science Foundation of China(61802436) and Natural National Key Basic Research Program of China(2016YFB050190104).
通信作者:王文娟(wwjhhx@sohu.com)

联，从而形成全局的攻击轨迹图。由于入侵检测系统(Intrusion Detection System，IDS)产生的每一条告警能够反映单步攻击的发生，将属于同一攻击行为的多个告警关联起来能够发现潜在的隐藏关系，还原完整的攻击场景。

本文在分析现有相关工作的基础上提出了一种基于因果知识和时空关联的攻击场景重构方法。首先，利用贝叶斯网络对因果知识进行建模，从具有 IP 地址相关性的告警序列中发掘出具有因果关系的攻击模式，为后续关联分析提供模板依据。然后，借助构建的因果知识网络，从因果、时间和空间等多个维度上对告警进行关联分析以发现潜在的隐藏关系。最后，通过实验验证了所提方法的有效性。本文所提方法能够还原攻击者完整的攻击渗透过程，重构出高层次的攻击场景，为构建可监管、可追责的云环境提供了一定的依据和参考。

2 相关工作

告警关联分析是重构攻击场景的主要途径之一。目前，告警关联分析所采用的技术手段主要包括以下 3 种：基于相似度的方法、基于数据挖掘的方法和基于因果知识的方法。

基于相似度的方法是通过计算告警之间各属性的相似度来判定告警之间是否具有关联关系，此种方法的关键在于定义相似度函数。Wang 等[6]定义了 3 种超告警类型，通过计算超告警间各属性相似度的加权平均和得到整体相似度，将整体相似度大于给定阈值的两个超告警进行关联。Mei 等[7]提出利用告警间的相似度函数对具有相似攻击行为的序列进行聚类，并基于聚类的告警序列自动发现多步攻击模式。此种方法计算开销小，且较少依赖专家知识，但需要预先定义阈值，只能发现统计上的关联情况，未能体现多步攻击步骤之间的因果联系，不易理解攻击过程和攻击意图。

基于数据挖掘的方法是通过关联算法、序列模式算法等发现隐藏在告警中的关联关系，然后依据关联关系构建攻击行为序列。Ge 等[8]利用 Map-Reduce 架构下的 Apriori 关联算法挖掘多维通信信息中的频繁项集，再进行综合关联分析。Lu 等[9]使用改进的 FP Growth 算法挖掘告警之间的关联规则，继而进行告警关联。该方法不依赖于专家知识，能够发现未知威胁行为，但是存在准确度和实用性较低等缺陷。

基于因果知识的方法认为多步攻击的各攻击步骤之间是存在因果关系的，通过制定因果关联知识来构建攻击场景。研究者从不同的角度或通过不同的方法来构建因果知识。

(1)谓词因果知识。该知识定义了每种攻击类型的前提条件集合和可能产生的后果集合。当告警 A 的后果能够完全或部分匹配于告警 B 的前提时，则 A，B 两个告警(A 先于 B 发生)能够关联。相关典型代表如 Steven 等领导的 JIGSAW 项目[10]和 Ning 等领导的 TIAA 项目[11]等。Zhang 等[12]提出基于攻击类型及其前提和条件建立攻击规划树，并提出基于攻击规划树的告警关联算法，实现实时告警关联及攻击场景重建。该方法存在的问题是：依赖于专家知识，一个新型的攻击无法与已知攻击关联，因为它的先决条件和后果是没有定义的。

(2)基于攻击图的因果知识。属性攻击图直观地呈现了攻击者利用各脆弱点进行逐步渗透的所有可能的攻击路径，描述了攻击步骤间的因果依赖关系。依据攻击图将告警进行匹配关联，从而构建攻击场景。因此，如何高效地生成攻击图，以及依据攻击图进行告警匹配关联成为该方法的两个重要研究内容。目前，一些研究更加专注于攻击图的分布式并行计算，以提高其生成效率。Wang 等[13]提出一种基于启发式搜索策略的全局攻击图生成方法，通过引入匹配索引表来存储原子攻击的最新匹配结果，从而提高攻击图的生成效率。Kaynar 等[14]提出了一种分布式攻击图生成算法，应用于分布式多代理平台上，由多个代理节点分别生成子攻击图，再由中心节点将子攻击图合并成全局攻击图。实验表明，分布式并行计算可以提高攻击图生成速度。然而，实际上，攻击步骤间的因果关系存在不确定性因素，且告警与攻击图的匹配度并不高，得到的结果不够完整。为了解决此问题，研究者将攻击场景构建视为不确定性推理问题，试图从概率推理的角度对攻击步骤之间的因果关系进行分析。

(3)基于概率推理的因果知识。Feng 等[15]基于马尔可夫性质挖掘出不同攻击类型间的转移概率矩阵，以构建攻击场景的因果知识，从而为告警关联提供模板依据。该方法能够自动地发掘出具有因果关系的攻击模式，但是生成的因果知识图中包含含圈路径，无法正确反映攻击类型间的因果关系和攻击过程的单调性，即攻击者不会再去获取已经具有的攻击能力。Liu 等[16]提出基于攻击图的多源告警关联分析算法，能够综合应用图关系和阈值进行告警的联动推断和预测，从而构建攻击场景。Lyu 等[17]深入分析了网络对抗的时空特性，用有限自动机模型模拟网络威胁渗透过程，用状态转移图挖掘威胁事件的时空关联关系，实时识别威胁状态。该方法采用自动机和状态攻击图，没有描述攻击状态间的不确定性转移关系，存在状态空间爆炸问题，难以适应大规模网络。

为了更直观地展示各类方法的特性，表 1 列出了不同方法的优缺点。

表 1 告警关联分析方法对比

Table 1 Comparison of alert correlation analysis methods

方法		优点	缺点
相似度		计算开销小，依赖专家知识少	需定义阈值，只能发现统计上的关联
数据挖掘		能够发现未知攻击	存在准确度、实用性较低等缺陷
因果知识	谓词	计算开销较小，准确度高	依赖专家知识，无法定义新型攻击的先决条件和后果
	攻击图	直观展示攻击步骤间的因果关系	状态空间爆炸，告警与攻击图的匹配度并不高
	概率推理	不受先验知识约束，可发现未知攻击，定量分析，准确度较高	没有描述攻击步骤之间的时间和空间关联关系

综上所述，基于概率的因果知识方法能够揭示告警之间的因果依赖关系以及概率推理关系，不受先验知识的约束，具有一定的未知攻击发现能力，准确度较高。该方法已成为目前攻击场景构建的主流方法。但是上述方法均没有描述攻击步骤之间的时间序列和空间关联关系，无法体现多步攻击随时间域和空间域同时变化的动态演化过程。因此，本文研究了如何从因果推理、时间和空间多维度上对告警进行关联分析，从而重构出高层次的攻击场景。

3 因果知识网络构建

我们在告警数据中建立因果知识网络，为了更好地说明本部分内容，首先给出告警的相关概念。

3.1 告警相关概念

定义 1（告警 alert） 是一个七元组 $a_i=\langle time, srcIP, dstIP, srcPort, dstPort, name, type\rangle$，其中 $time$ 是时间戳，为传感器检测到恶意攻击时所产生告警的时间；$srcIP$ 和 $dstIP$ 是源 IP 地址和目的 IP 地址；$srcPort$ 和 $dstPort$ 是源端口和目的端口；$name$ 表示告警名称，表明发生了什么攻击；$type$ 表示告警的攻击类型，包括 4 大类攻击类型，即探测类（Probe）、远程访问类（R2L）、提升权限类（U2R）和拒绝服务类（DoS）等。

攻击者对网络的入侵一般有一个攻击周期，若其在长时间内仍未发起后续攻击，则认为攻击者入侵失败，设置一个时间窗口来衡量攻击者的成功与否。

定义 2（时间窗口 ΔT） 已知大部分攻击的一个攻击周期为 $2h$，因此设置 $\Delta T=2\text{h}$。

IDS 产生的告警不可避免地存在一些重复告警，重复告警主要是由同一攻击源用同一种攻击方式对目的主机进行多次不同时间的攻击所造成的，因此有必要消除重复告警，将其简化合并为同一条告警。此外，由于同一攻击活动触发的告警事件，彼此在 IP 地址分布上总是具有关联性，依据这种 IP 地址相关性，将同一攻击活动的告警事件聚合在一起，能够为后续因果知识构建提供更精简、准确的告警数据，从而避免关联关系混乱。

定义 3（重复告警） 如果两条告警 a_i 和 a_j 的 $\langle srcIP, dstIP, srcPort, dstPort, name\rangle$ 字段都相等且在一定时间周期内，那么认为 a_i 和 a_j 具有重复关系，应该合并。

定义 4（IP 地址相关性） 如果告警 a_i 的 IP 地址无论是源 IP 地址或目的 IP 地址，总有一个和 a_j 相同，且 a_i 和 a_j 在一定时间周期内，那么这两条报警是 IP 地址相关的。

定义 5（告警类簇，Alert Cluster，AC） 指将具有 IP 地址相关性的告警按时间戳进行排列所构成的序列，记为 $AC=\{a_1,\cdots,a_n\}$，满足 $a_i.time<a_j.time$ $(1<i<j<n)$。

基于 IP 地址相关性的告警聚类过程采用文献[15]所提的方法，这里不再论述。将处理后生成的类簇 AC 作为数据源，用来建立因果知识网络，从而发现 AC 中各告警名称间的因果关联关系，表 2 列出了告警类簇实例。

表 2 告警类簇实例

Table 2 Instance of alert cluster

T	$srcIP$	$dstIP$	$name$	$type$	pri
t_1	IP_1	IP_2	a	Probe	1
t_2	IP_1	IP_2	b	R2L	2
t_3	IP_1	IP_3	a	probe	1
t_4	IP_1	IP_3	b	R2L	2
t_5	IP_2	IP_5	c	U2R	3
t_6	IP_3	IP_5	c	U2R	3

3.2 因果知识网络定义

攻击步骤之间存在不确定性的因果关系，由于贝叶斯网络能够表示随机变量间的因果关系及关系依赖程度，具有很强的概率推理能力，对未学习的知识也具有很强的预测能力[18]，故采用贝叶斯网络来构建因果知识网络。

定义 6（因果知识网络，Causal Knowledge Network，CKN）由网络结构和网络参数两部分组成，记为 $CKN=(G,\Theta)$。

（1）网络结构 G：$G=(V,E)$，G 是一个有向无环图（Directed Acyclic Graph，DAG）。其中：

1）V 为节点集合，变量 $v_i\in V$ 表示具体攻击名称，表明攻击者发动了什么攻击，其取值为“True”或“False”，对应于节点的状态（State），表示该攻击是否发生。

2）E 为有向边集合，每条边反映节点之间存在的因果依赖关系。如果存在 $v_i\rightarrow v_j$，则称 v_i 是 v_j 的父节点，v_j 是 v_i 的子节点，表示只有攻击步骤 v_i 发生了，攻击步骤 v_j 才有可能发生，用 $Pa(v_i)$，$Ch(v_i)$ 分别表示 v_i 的父节点和子节点集合。

（2）网络参数 Θ 为一组局部条件概率分布表（Conditional Probability Table，CPT）。$\theta_i\in\Theta$ 表示节点 v_i 的局部条件概率 CPT，$V=\{v_1,\cdots,v_n\}$ 的全局联合概率 P 为：

$$P(V)=P(v_1,v_2,\cdots,v_n)=\prod_{i=1}^{n}P(v_i\mid v_{i-1},\cdots,v_1)$$

全局联合概率表示为每个节点的局部条件概率的乘积。

因果知识网络具有以下性质。

性质 1（条件独立） v_i，v_j 和 v_k 是有向图 G 中的任意节点，若满足 $P(v_i\mid v_j,v_k)=P(v_i\mid v_k)$，则称 v_i 和 v_j 关于 v_k 条件独立，记作 $v_i\perp v_j\mid v_k$。

性质 2（局部马尔可夫） 在给定节点 v_i 的父节点条件下，节点 v_i 条件独立于其非子节点。

依据马尔可夫性质，联合概率 P 可以表示为每个节点关于其父节点的条件概率的乘积，即：

$$P(V)=P(v_1,v_2,\cdots,v_n)=\prod_{i=1}^{n}P(v_i\mid Pa(v_i))$$

其中，$Pa(v_i)$ 表示节点 v_i 的父节点集合。节点 v_i 的 CPT 为 $P(v_i\mid Pa(v_i))$。

图 1 所示为因果知识网络。其中包括 4 个变量的因果知识网络。例如，攻击者利用 *nmap* 漏洞扫描工具能够发现服务存在的漏洞，利用漏洞有可能进一步实施 *buffer_overflow* 缓冲区溢出攻击。同样，攻击者进行 *portsweep* 端口扫描也能够探测服务，进一步开展 *buffer_overflow* 攻击。图中展示了节点 *buffer_overflow* 的 CPT，CPT 中每个条目表示 buffer_overflow 的每个已知值的条件概率，对应于其父节点的每个可能的组合。图 1 的联合概率为：

$$P(v_1,v_2,v_3,v_4)=P(v_1)P(v_2)P(v_3\mid v_1,v_2)P(v_4\mid v_3)$$

其中，存在序列连接 $v_1\rightarrow v_3\rightarrow v_4$ 和收敛连接 $v_1\rightarrow v_3\leftarrow v_2$，在给定 v_3 条件下，v_1 和 v_4 关于 v_3 条件独立，而 v_1 和 v_2 之间是存在相关性的。

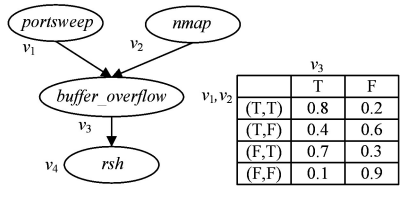

图 1 因果知识网络

Fig. 1 Causal knowledge network

不难发现，在因果知识网络中，网络结构 G 能够以有向无环图的形式直观地表示攻击节点之间的因果关系，是定性知识的表示部分；网络参数 Θ 则能够通过计算局部条件概率

来量化每个节点对其父节点的依赖关系的强度，是定量知识的表示部分。通过定性与定量的结合，有利于准确、合理地构建攻击节点的因果知识网络。

3.3 因果知识网络的构建

因果知识网络构建包括网络结构和网络参数确定两个部分。因果知识网络的构建思想是利用局部马尔可夫性质进行条件独立性检验，判断节点间的独立性和相关性，并根据独立性或相关性构建出相应的有向无环图结构。构建因果知识网络过程如下。

首先，从告警类簇 AC 中依次提取告警名称，形成节点集 $V=\{v_1,\cdots,v_n \mid v_i=a_i.name, v_i\neq v_j\}$。

其次，对于 $i=1,\cdots,n$，依次在网络中添加 v_i 节点，并计算 v_i 的局部条件概率。对于 v_i 节点，判断在节点集合 $\{v_{i-1},\cdots,v_1\}$ 中是否存在节点子集 X，使得 $P(v_i \mid v_{i-1},\cdots,v_1)=P(v_i \mid X)$ 成立，如果存在，则子集 X 为节点 v_i 的父节点集，即 $X=Pa(v_i)$。在搜索节点子集 X 时，应首先考虑节点数目为 1 的子集来判断是否条件独立，然后考虑节点数目为 2，之后不断增加节点数目直至等式成立。

最后，综合每个节点形成因果知识网络。这里，$P(v_i \mid v_{i-1},\cdots,v_1)$ 表示告警序列中出现告警名称 $v_1,\cdots,v_{i-1}$ 后，接下来告警名称为 v_i 的条件概率，即实施攻击步骤 $v_1,\cdots,v_{i-1}$ 后，攻击者下一步的攻击步骤为 v_i 的条件概率。通过计算 $P(v_i \mid Pa(v_i))$，可得到因果知识网络中节点 v_i 的网络参数。挖掘算法具体如算法 1 所示。

算法 1 挖掘因果知识过程

输入：告警类簇 AC=$\{a_1,\cdots,a_n\}$

输出：因果知识网络 CKN

```
1. Creat the set of attack name V;
2. Foreach a_i:AC
3.   {if i==1
4.       v_i=a_i.name and v_j=a_{i+1}.name;
5.       add v_i and v_j to the set V;
6.       num_ij++; //用于统计攻击名称 v_i后面直接跟 v_j的个数
7.       P(v_j|v_i)=num/sum
8.    else
9.       v_i=a_i.name and v_j=a_{i+1}.name;
10.      if v_j is not included in V
11.         add v_j to V;
12.         num_ij++;
13.      else
14.         num_ij++;}
15. m=num{V}; //m 为 V 中元素的个数
16. For(i=2;i<m;i++)
17.   {search X let X⊂(v_{i-1},...,v_1)
18.      if exist P(v_i|v_{i-1},...,v_1)=P(v_i|X)
19.         X=Pa(v_i);
20.      else
21.            v_i=root; //v_i为根节点;
22. i++; }
23. end.
```

4 基于时空关联的攻击场景构建

4.1 时空关联分析思想

因果知识网络建立后，接下来从时间和空间两个维度对告警序列进行关联分析，从而发现不同时间段内不同位置攻击事件潜在的隐藏关系：1）通过时间序列分析，能够体现多步攻击的渗透过程及其攻击行为模式的规律性；2）通过空间关联分析，能够体现网络攻击行为在地址分布/攻击位置上的关联性，从而追踪到攻击源。

告警时空关联分析的思想如图 2 所示，设置滑动关联时间窗口 $T=\Delta T$。首先，挖掘出同一时间窗口内告警之间的空间关联。网络攻击行为是受网络拓扑约束的，同一攻击活动触发的告警在地址分布上总是具有关联性。例如多步攻击中，前一攻击步骤的目标节点可能就是下一攻击步骤的源节点。因此，通过 IP 地址相关性分析能够从空间维度上发现告警事件潜在的隐藏关系。其次，具有地址相关性的告警之间是受前因后果关系约束的，也就是说，前一个攻击步骤为后续步骤提供了前提条件，可以依据已建立的因果知识进行约束。即如果两个告警事件之间是 IP 地址相关的，但是攻击名称之间不存在因果关系，则不符合攻击行为模式的规律性，这两个告警不应关联。此外，不同时间窗口之间的告警也可能存在关联性，可以通过关联多个时间段的场景发现。

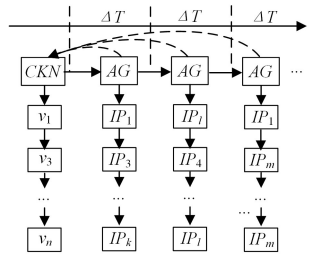

图 2 告警时空关联关系

Fig. 2 Alert temporal-spatial correlation

4.2 时空攻击场景构建

借鉴脆弱性攻击图的分析方法，从告警的角度构建基于时空关联分析的攻击场景图。

定义 7（时空攻击场景 AG） 描述告警集合 AS 中反映各主机节点间时空关联关系的攻击序列，可以形式化描述为一个三元组 $AG=(V,E,Na)$。其中：

（1）V 是主机节点集合，$V=\{v_i \mid v_i=IP_1,\cdots,IP_m\}$，告警中的 $srcIP$ 和 $dstIP$ 地址表示某次攻击的攻击者所在主机和攻击目标所在主机。

（2）E 为有向边集合。假定存在攻击节点 v_i 和被攻击节点 v_j，则有 $E=\{e_{ij} \mid e_{ij}=v_i\times v_j\}$。每个节点都有其入边和出边，入边代表从其他节点（攻击）到自身，出边代表从自身（攻击）到其他节点。

（3）a 为依附在有向边上的攻击名称，描述了该有向边所代表的具体的攻击行为。针对某节点 v_i，a_i^{pre} 表示节点 v_i 入边上的前件攻击，a_i^{post} 表示节点 v_i 出边上的后件攻击。

对于节点 v_i，其入边和出边之间存在以下 3 种情况。

（1）如果前件攻击 a_i^{pre} 与后件攻击 a_i^{post} 之间不存在因果关系，则 v_i 的入边与出边串联失败。这意味着即使前一攻击步骤的目的地址是下一攻击步骤的源地址，但是攻击名称间不存在因果关系，因此不能够级联，有可能发生了漏报、误报或乱序，如图 3(a)所示，虚线表示级联失败。

（2）如果前件攻击 a_i^{pre} 与后件攻击 a_i^{post} 之间存在因果关

系，则入边与出边串联成功。这意味着前件攻击的目标地址就是后件攻击的源地址，且前件攻击成功实施导致其后果被满足。如图 3(b)所示，实线表示串联成功。

(3)当节点 v_h 对节点 v_i 产生多次不同类型的攻击行为，且攻击名称之间存在因果关系时，攻击行为之间存在串联关系；而当攻击名称之间不存在因果关系时，攻击行为之间存在并联关系，如图 3(c)所示。

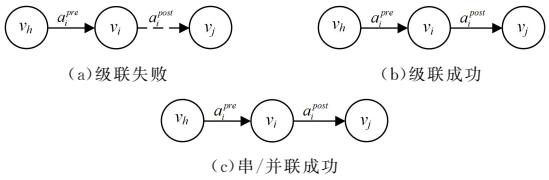

图 3 告警攻击场景构建过程

Fig. 3 Alert attack scenario construction process

构造攻击场景的过程实际上是有向图的邻接表的建立过程。每一个主机节点作为攻击场景中的一个顶点 v_i，与之相邻的所有顶点存放在一个链表中，即为顶点 v_i 的邻接表。邻接表除了存储与该顶点相邻的顶点，还要存储相关的攻击类型。在遍历有向图的顶点集时，可以采用深度优先遍历算法，从某个顶点 v 出发，找到一个与 v 相邻且没有被访问过的顶点 w。然后从 w 开始再进行深度优先遍历，这时顶点 w 的入边和出边是否能够级联，需要依据已建立的因果知识进行判断。此过程依次类推，直到所有顶点全部被遍历。图 4 展示了表 1 所对应的告警邻接表，图 5 展示了其所对应的攻击场景图。前提是攻击名称 a 与 b、b 与 c 之间存在因果推理关系。

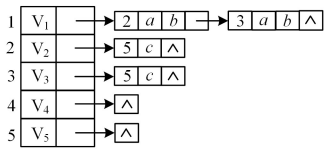

图 4 告警邻接表

Fig. 4 Alert adjacency list

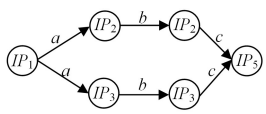

图 5 攻击场景示例

Fig. 5 Example of attack scenario

5 实验测试分析

5.1 实验环境

实验环境搭建借鉴了 Cui 等[19]的思想，如图 6 所示。云虚拟化平台采用 Xen 系统，云物理机(Physical Machine，PM)包括特权域 dom0 和非特权域 domU，domU 由若干个虚拟机(Virtual Machine，VM)组成。云物理机之间由传统交换机连接，服务器内 VM 之间由可编程虚拟交换机(Open vSwitches，OVS)连接。由于 OVS 仅完成数据转发这一功能，而路由控制则由网络控制器(Network Controller，NC)来完成。NC 和 OVS 模块将虚拟网络流重定向到 NIDS 系统中进行入侵检测。NIDS 作为一个虚拟应用，部署在非特权域 domU 中，使得其动态迁移到任意网段中会变得相对容易，且减轻了特权域的负载。设置 1 台 VM 为攻击方，1 台 VM 为攻击目标，攻击方 VM OS 为 Linux 系统，攻击目标 VM OS 为 win7 系统。攻击方利用 Tcpreplay 工具向攻击目标重放 DARPA 2000 LLDOS 1.0 数据集，该数据集是目前最全面的攻击测试数据集。Snort-IDS 系统检测到该攻击数据并产生告警，告警名称及编号如表 3 所列。

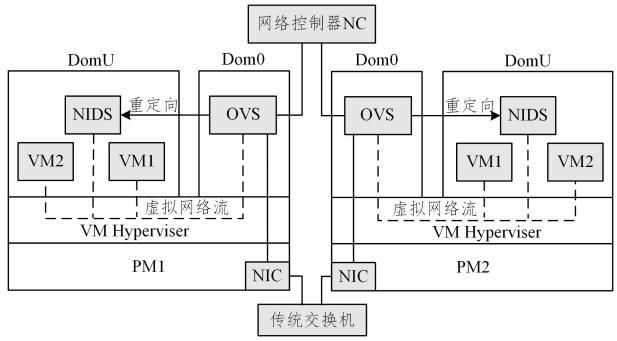

图 6 Xen 云平台环境

Fig. 6 Xen cloud platform environment

表 3 攻击名称及其编号

Table 3 Attack names and their numbers

Num	Alert-name	Type
att_1	ICMP PING	probe
att_2	RPC portmap sadmind request UDP	probe
att_3	RPC sadmind UDP PING	probe
att_4	RPC sadmind query with root credentials attempt UDP	R2L
att_5	RPC sadmind UDP NETMGT_PROC_SERVICE CLIENT_DOMAIN overflow attempt	U2R
att_6	RSERVICES rsh root	U2R
att_7	MStream Command	U2R
att_8	BAD-TRAFFIC loopback traffic	DOS

5.2 实验结果

生成具有 IP 地址相关性的告警类簇之后，按照算法 1 构建因果知识网络，整个因果知识网络包含 8 个攻击名称，图 7 中有向边即表示告警名称之间存在因果关系，这里省略了每个告警名称节点的条件概率表。

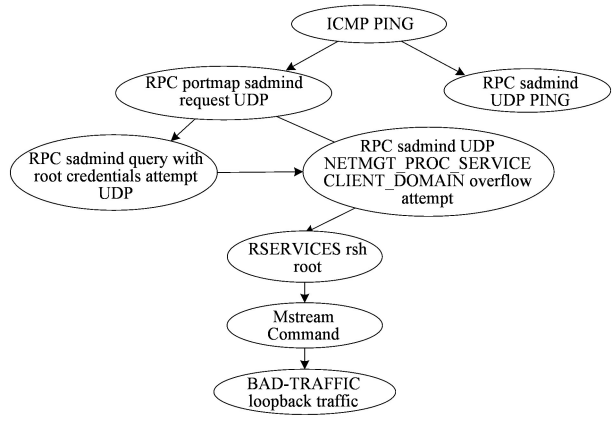

图 7 DDOS 攻击因果知识网络

Fig. 7 Causal knowledge network of DDOS attacks

告警事件之间是存在关联性的，依据已建立的因果知识网络，建立基于时空关联的攻击场景。上述生成的具有 IP 地址相关性的告警类簇中，最后一条告警与第一条告警的时间间隔为 36 min，如果设置的关联时间窗口 ΔT 较小，假定时间窗口 $\Delta T=6$ min，则能够观察到 DDOS 攻击的逐步渗透过程。DDOS 时空关联图如图 8 所示，用不同的颜色表示不同的攻击类型。黑色表示 Probe 攻击，橙色表示 R2L 攻击，紫色表示 U2R 攻击，红色表示 DoS 攻击。第 1 步，攻击者从源 IP 地址 202.77.162.213 对 3 个目标网络 172.16.112.0/24，172.16.113.0/24，172.16.115.0/24 等发送 ICMP echo 请求，以确定主机开放情况，通过扫描发现多台存活主机。第 2 步，攻击者向存活主机发送 RPC 请求，询问 sadmind 服务，查询发现有 3 台主机运行 sadmind 服务并回复了 RPC 请求，这 3 台主机便作为攻击者的傀儡机，用黑色表示主机/端口扫描攻击。第 3 步，攻击者尝试远程连接傀儡机失败后，进行 sadmind 缓冲区溢出攻击，获得傀儡机的 root 权限，这里用橙色表示 R2L 攻击，紫色表示 U2R 攻击。第 4 步，攻击者以 root 权限登录傀儡机并在傀儡机上安装 mstream 守护程序，准备进行分布式拒绝服务攻击。第 5 步，攻击者操控傀儡机通过伪造大量随机 IP 地址（伪造的地址都是 127 的回环地址），向攻击目标 131.84.1.31 发送大量的 TCP 数据包，以进行洪泛攻击，这里用红色表示 DOS 攻击。

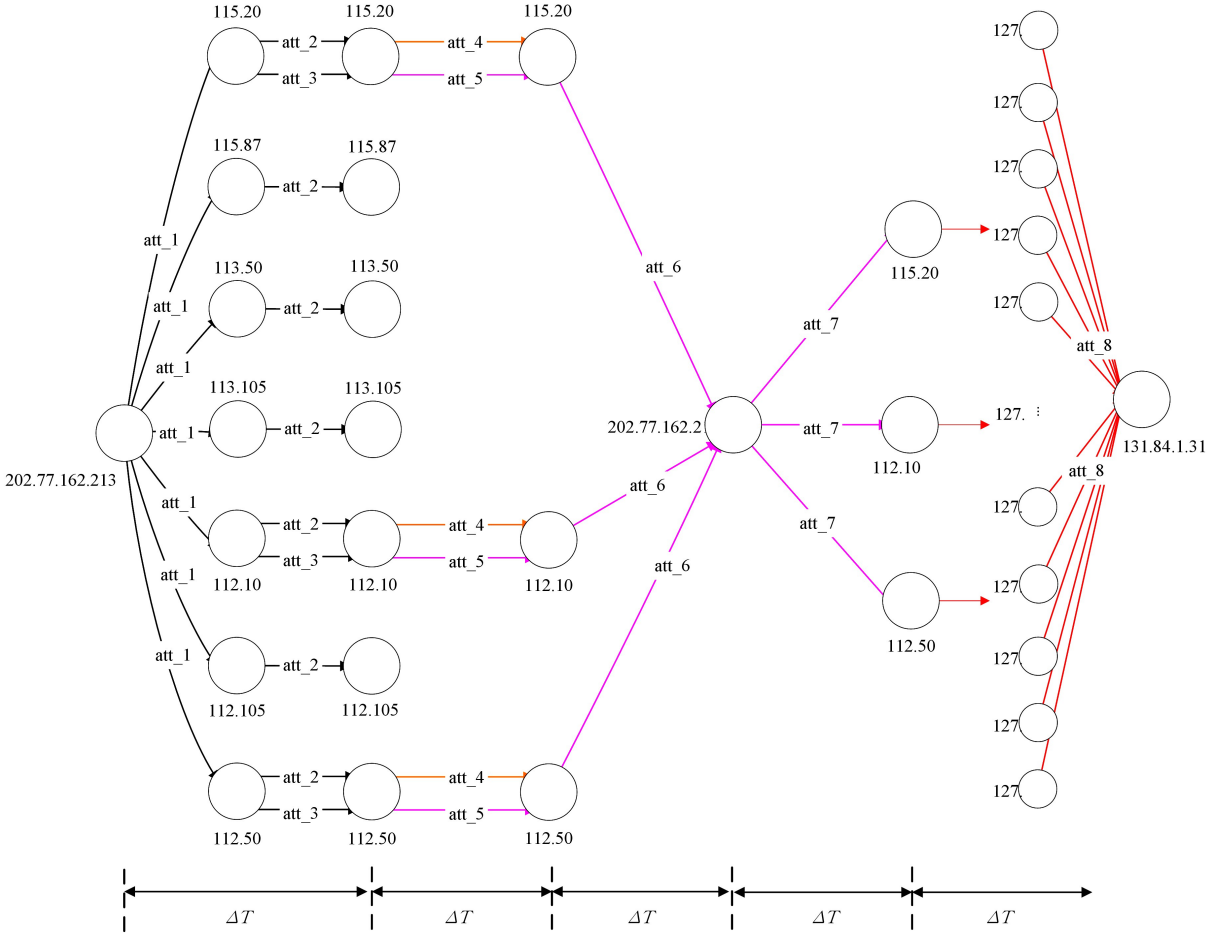

图 8 DDOS 攻击时空关联图（电子版为彩色）

Fig. 8 Temporal-spatial correlation graph of DDOS attacks

5.3 实验结论分析

利用贝叶斯网络构建因果知识，能够获取定性的因果关系与定量的因果依赖程度，具有良好的知识表示能力。贝叶斯网络具有条件独立与局部马尔可夫性质，在计算某些攻击名称节点的后验概率时可以减少计算量，大大降低因果知识挖掘的复杂度，保证因果知识网络的高效生成。构建攻击场景的过程是告警邻接表及图的建立过程。假设 AC 的告警数量为 N，则遍历整个类簇构建邻接表的复杂度为 $O(N)$。建立场景图时需要遍历查询图中的顶点集，最多 $2N$ 个顶点，则整个算法复杂度为 $O(N^2)$。

结束语 安全问题已经成为制约云计算推广和发展的主要障碍，云环境下的攻击行为表现出隐蔽性好、复杂多步等特点。揭示了攻击步骤之间隐藏的逻辑关系、重构出攻击场景是云计算亟需解决的难题之一。本文提出了基于因果知识和时空关联的攻击场景重构技术。首先，利用贝叶斯网络对因果知识进行建模，从具有 IP 地址相关性的告警序列中发掘出具有因果关系的攻击模式，为后续关联分析提供模板依据。然后，借助构建的因果知识网络，从因果、时间和空间多维度上对告警进行关联分析以发现潜在的隐藏关系。最后，通过实验验证了所提方法的有效性。本文所提方法能够还原攻击者完整的攻击渗透过程，重构出高层次的攻击场景，为构建可监管、可追责的云环境提供了一定的依据和参考。

参 考 文 献

[1] PETER M M, TIMOTHY G. The NIST Definition of Cloud Computing[M]. National Institute of Standard & Technology, 2011.

[2] The Notorious Nine: Cloud Computing Top Threats in 2013 [EB/OL]. http://www.cloudsecurityalliance.org/group/top-threats.

[3] CHEN X J, FANG B X, TAN Q F. Inferring attack intent of malicious insider based on probabilistic attack graph model[J]. Chinese Journal of Computer, 2014, 34(1): 62-72.

[4] WANG L. Study on Method of network multi-stage attack plan recognition[D]. Wuhan: Huazhong University of Science and Technology, 2007.

[5] PENG N, YUN C, DOUGLAS S. R Constructing attack scenarios through correlation of intrusion alerts[C]// ACM Symposium on Computer and Communications Security. Washington, DC, United States, 2002: 245-254.

[6] WANG L, GHORBANI A A, LI Y. Automatic multi-step attack pattern discovering[J]. International Journal of Network Security, 2010, 10(2): 142-152.

[7] MEI H B, GONG J, ZHANG M H. Research on discovering multi-step attack patterns based on clustering IDS alert sequences[J]. Journal on Communications, 2011, 32(5): 63-69.

[8] GE L, JI X S, JIANG T. Association rules and its implementation in Map-Reduce[J]. Journal of Electronics & Information Technology, 2014, 36(08): 1831-1837.

[9] LU X G, DU X H, WANG W J. Alert correlation algorithm based on improved FP growth[J]. Computer Science, 2019, 46(8): 64-70.

[10] STEVEN J T, KARL L. A requires/provides model for computer attacks[C]// Proc. of the 2000 Workshop on New Security Paradigms. New York: ACM, 2000: 256-263.

[11] NING P. TIAA: A visual toolkit for intrusion alert analysis [M]. North Carolina State University at Raleigh, 2003.

[12] ZHANG J, LI X P, WANG H J. Real-time alert correlation approach based on attack planning graph[J]. Journal of Computer Applications, 2016(6): 1538-1543.

[13] WANG S, TANG G, KOU G. An attack graph generation method based on heuristic searching strategy [C] // IEEE International Conference on Computer & Communications. IEEE, 2017.

[14] KAYNAR K, SIVRIKAYA F. Distributed attack graph generation[J]. IEEE Transactions on Dependable and Secure Computing, 2016, 13(5): 519-532.

[15] FENG X W, WANG D X, HUANG M H. A Mining Approach for Causal Knowledge in Alert Correlating Based on the Markov Property[J]. Journal of Computer Research and Development, 2014, 51(11): 2493-2504.

[16] LIU W X, ZENG K F, WU B. Alert processing based on attack graph and multi-source analyzing[J]. Journal on Communications, 2015, 36(9): 135-144.

[17] LYU H Y, PENG W, WANG R M. A Real-time Network Threat Recognition and Assessment Method based on Association Analysis of Time and Space[J]. Journal of Computer Research and Development, 2014, 51(5): 1039-1049.

[18] XIE P, LI J H, OU X, et al. Using bayesian networks for cyber security analysis[C]// Proceedings of the 2010 IEEE/IFIP International Conference on Dependable Systems and Networks. Chicago, IL, USA, IEEE, 2010.

[19] CUI J S, GUO C, CHEN L. Establishing process-level defense-in-depth framework for software defined networks[J]. Journal of Software, 2014, 25(10): 2251-2265.

WANG Wen-juan, born in 1981, postgraduate, associate professor. Her main research interests include information security and cloud computing.