反应式容侵系统入侵预测的混合式贝叶斯网络方法

王良民[1,2], 马建峰[3]

(1. 东南大学计算机科学与工程学院, 南京 210018; 2. 江苏大学计算机科学与通信工程学院, 江苏镇江 212013;
3. 西安电子科技大学计算机网络与信息安全教育部重点实验室, 西安 710071)

[摘要] 为解决反应式容忍入侵系统中的入侵预测问题,提出了新的混合式贝叶斯网络方法。该方法中,提出了一种基于系统安全状态的入侵模型,以攻击者能力上升的过程来描述入侵,关注入侵对系统的影响,适合于反应式容侵系统根据当前状态选择合适的响应机制。提出了基于入侵模型的混合式贝叶斯网络(HYBN; hybrid bayesian network)模型,将入侵模型中攻击行为和系统安全状态节点分离为攻击层和状态层两个网络层次,两层间使用收敛连接,而两层内部的节点间使用连续连接。在特定的信度更新算法的支持下,实验说明该贝叶斯网络方法用于入侵预测的有效性,比较说明 HYBN方法的优点。

[关键词] 容忍入侵; 警报关联; 入侵模型; 入侵预测

[中图分类号] TP393 [文献标识码] A [文章编号] 1009-1742(2008)08-0087-10

1 前言

容忍入侵技术[1,2](intrusion tolerance)能在故障/入侵已发生时保证系统关键功能继续执行,受到了越来越多的关注,其中包括美国的 OASIS和欧洲的 MAFTIA这两个重要的研究计划。在这两个计划的支持和带动下,有关容忍入侵技术的研究得到了迅速发展。容忍入侵技术主要分为两类,一是先应式(proactive)的,另一是反应式(reactive)的。先应式容侵技术假设计算环境是不可信的,从开始就重新设计整个系统结构,使可信的部分系统能够在不可信的环境中安全地合作,以保证攻击发生后对系统没有太大的影响。目前先应式的容侵技术多以门槛密码技术、Byzantine协议技术等为理论基础[3,4]。反应式容侵技术在检测到局部系统失效或估计到系统被攻击后,调整系统结构,重新分配资源,从而达到继续服务的目的。基于反应式容侵技术的系统一般都包括一个基于风险概念的入侵预测系统、系统资源控制系统和在线的修复管理程序,比较典型的反应式容侵系统有 ITDB数据库系统[5]、Internet的服务保护系统[6]等。先应式容侵系统不用考虑入侵,但是需要重新设计系统,难以和现有的设备与系统兼容;反应式容侵系统不需要重新设计系统结构,系统的操作和连接界面也可以与原有系统保持一致,受到具体应用场合的欢迎,但是它依赖于由入侵预测技术构成的容忍入侵触发器。

对于入侵预测技术的研究,面临着两方面的困难:一是入侵模型问题,由于入侵是恶意对手的蓄意行为,它不同于符合随机分布的故障,很难描述和预测,文献[1]将入侵模型的描述与构建列为一个公开问题;二是预测方法问题,通常入侵预测的依据是来自入侵检测系统提供的警报,而当前的入侵检测系统拥有太高的误警率和漏警率[7],对入侵预测方法提出了更高要求,需其能在一定程度上容忍误报、

[收稿日期] 2007-05-23; 修回日期 2007-09-10

[基金项目] 国家自然科学基金资助项目(60703115, 60503012, 90604003); 江苏省自然科学基金资助项目(BK2007560, BK2007708); 国家博士后科学基金(20070420955); 江苏省博士后科研资助计划(0702003B); 江苏大学高级人才科研启动经费(07JDG080)

[作者简介] 王良民(1977-), 男, 安徽潜山县人, 博士,江苏大学讲师,东南大学博士后,主要研究领域为网络安全结构,容忍入侵理论与方法等; 马建峰(1963-), 男,陕西临潼县人,博士,西安电子科技大学教授,博导, 主要研究领域为信息安全与密码学

漏报并根据推理给出正确的预测。

2 入侵模型及其构建方法

入侵是恶意对手的蓄意行为, 不符合随机分布规律, 基于随机分布的故障模型并不适用, 使入侵模型的描述与构建成为难题[1]。反应式容侵系统的入侵模型, 应立足于容侵系统本身, 把攻击者发起的不断升级的入侵行为看成系统安全状态下降的过程, 而系统安全状态的下降, 是通过攻击者针对系统操控能力的提高来体现。为此, 提出了一种基于攻击者能力的入侵模型, 该入侵模型通过攻击行为内在的逻辑联系, 以攻击者能力提升的过程来描述入侵及入侵导致的系统安全状态变迁, 使系统能及时了解自身安全状态下降, 当其下降到某一阈值时, 可以方便地启动相应的容侵机制。

2.1 基于攻击者能力的入侵模型

文献[8, 9]从攻击者的角度出发, 认为入侵是这些有关联的攻击步骤。笔者立足于被破坏系统, 认为入侵是攻击者获取了针对该系统的操控能力, 并根据攻击者能力不断提升(即系统安全性逐渐降低)的过程选用文献[10]关于攻击者能力的描述来建立入侵模型。

定义 1 攻击者所具有的针对计算机系统的操控能力(capability)是一个 6元组:

$$\text{Capability} = (\text{Source}, \text{Target}, \text{Action}, \text{Service}, \text{Property}, \text{Credential}) \quad (1)$$

Capability描述了攻击者的行为能力, 表示来自 Source攻击者, 用证书 Credential, 针对 Target中 Service服务的 Property属性, 采取 Action行为。

例如 $C_1 = (\text{src}, \text{tgt}, \text{read}, \text{passwd}, \text{content}, \text{smith})$, 表示来自 src的攻击者, 他以 smith的证书获得了针对目标 dst的 content服务中 passwd文件的 read权限。此外, 不同的能力之间可能具有某种逻辑联系。如能力 $C_2 = (\text{src}, \text{tgt}, \text{read}, \text{All files}, \text{content}, \{\text{smith}, \text{john}\})$, 显然 C_2的能力比 C_1表示的能力大, 或 C_2蕴涵了 C_1。

定义 2 C_1和 C_2是 2个不同的攻击能力, 如果它们具有如下关系:

1) 相同的 Source, Target和 Action;

2) 相同的类型 Service和 Property;

3) C_1的 Service, Property和 Credential分别是 C_2的 Service, Property和 Credential的子集。

则称 C_2蕴涵 C_1, 或者 C_1可由 C_2推出, 记为 $C_2 \Rightarrow C_1$。

由于入侵是系统安全性不断下降即攻击者能力不断提升的过程, 可以用图 1的基于攻击者能力的状态转移过程作为入侵模型, 图 1中的节点 S是攻击者的能力的集合, 边 a表示攻击。在初始状态 $S_0 = \{\}$, 攻击者发起攻击行为 a_1, 获得能力 C_1, 进入状态 $S_1 = \{C_1\}$; 在具备了发起攻击 a_3的准备条件 C_1之后, 实施 a_3攻击, 获取新的能力 C_3, 进入状态 $S_3 = \{C_1, C_3\}$; 而在通过 a_2获得能力 C_2之后, S_2和 S_3的并集 $\{C_1, C_2, C_3\}$构成了发动 a_4攻击的准备条件, 发动 a_4攻击, 获取新能力 C_4状态转移到 $S_4 = \{C_1, C_2, C_3, C_4\}$, ……

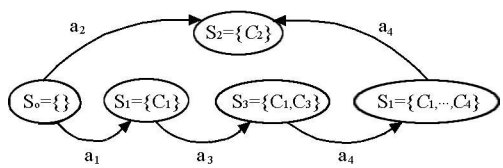

图 1 基于攻击者能力的入侵模型

Fig. 1 Intrusion model based abilities of attackers

定义 3 基于攻击者能力的入侵模型是一个状态转移图, 图 1中状态节点是攻击者能力的集合, 攻击者能力由定义 1给出; 图 1中的边由导致状态转移的攻击行为或其触发的警报组成。

由图 1可知, 任何一个攻击行为的发起, 都需要一定攻击能力作为准备条件, 同时, 该攻击行为发生以后, 又获得了新的攻击能力, 给后续的攻击提供准备。也就是说, 任何一个攻击行为必须与两个能力状态相对应, 一个是发起该攻击的初始状态, 是该攻击能成功发起的先决条件; 另一个是该攻击奏效后的结果状态, 包含了该攻击所获取的新能力。为此, 用超警报来描述警报与这两个能力的关联。

定义 4 一个超警报(hyper_alert)由 3部分组成:

$$\text{Hyper\_alert} = (\text{Alert}, \text{Prerequisite}, \text{Consequence}) \quad (2)$$

其中, Alert表示该攻击发生时的 IDS警报, 由一个 4元组构成, $\text{Alert} = (\text{name}, \text{time}, \text{source}, \text{target})$, 它的组元 name是警报名称; time本身也是一个 2元组 $\text{time} = (\text{begin}, \text{end})$, 分别表示该警报所表示攻击行为的发起时间和结束时间; 组元 source, target分别表示攻击行为的来源和攻击目标。Prerequisite表示 Alert产生之前必须具备的先决条件, 是能力的合取; Consequence是 Alert所表示的攻击奏效后, 攻击

者具有的能力的合取。

超警报的定义包含了警报发生前必须具备的条件和将产生的后果,给后续处理提供了有力帮助。

根据超警报和入侵模型的定义可知,由于攻击者在当前能力状态的基础上发动攻击,攻击行为成功后,下一个状态中的能力显然包含前一个状态中的所有能力。

引理 1 超警报 Hyper_alert=(Alert, Prerequisite, Consequence)具有 Consequence⇒Prerequisite。

定义 1 以基于攻击者能力的状态转移图,形象具体地描述了一个入侵事件,如果将该入侵事件中警报和攻击事件看成变量,则该状态转移图描述了一个入侵行为模式。在已知此入侵模式的情况下,可以用超警报来描述这个模式。但是,没有先验的入侵模型,只能根据 IDS提供的警报流,发现入侵事件并自适应模仿攻击者的逻辑构建入侵模型。

2.2 入侵模型与元攻击及其构建方法

为发现入侵模式,将 IDS提供的警报流转化为超警报流,并从中寻找这些超警报间的逻辑关联。

定义 5 一个元攻击(Meta_attack)是一个 3元组

$$Meta\_attack=\{H\_set, C\_set, time\} \quad (3)$$

其中 H_set是关联在一起的超警报集合, C_set是集合 H_set中所有超警报获取的能力的集合, time=(begin, end)是一个 2元组, begin是 H_set中所有超警报开始时间的最小值, end是 H_set超警报结束时间中的最大值。

元攻击的定义已经将图 1所示的攻击模型中节点集(能力状态)和边集(警报)以及节点与边之间的对应关系做了描述,因此,由 IDS警报构建入侵模型的问题,可以转化为如何由超警报流构建元攻击的问题。

定义 6 2个超警报 $h_i=(A_i, Pre_i, Cons_i)$ 和 $h'=(A', Pre', Cons')$之间满足

$$A_i\_Time\_end<A'\_Time\_begin \quad (4)$$

如果 $Pre'\Rightarrow Cons_i$则称 h_i是 $H a'$的必要准备,简称为准备,记为 h_i Prep for h';当超警报集合 $\{h_i \mid h_i=(A_i, Pre_i, Cons_i), i=1, 2, \cdots, n\}$满足

$$Cons_1\wedge Cons_2\wedge\cdots\wedge Cons_n\Rightarrow Pre' \quad (5)$$

$$Cons_1\wedge Cons_2\wedge\cdots\wedge Cons_n\Rightarrow\text{不能蕴涵 } Cons' \quad (6)$$

则称 $\{h_i \mid i=1, 2, \cdots, n\}$中 n个超警报构成 h的充分准备。

在 IDS提供的警报中,由于多传感器的应用,常常出现针对同一个攻击的多个报警,为此需要进行警报融合和聚类。

假设 1 任何一个入侵模型中,入侵者都不会采取重复的多样性的行动去获取已有的攻击能力。

对同一个攻击者来说,当他已经具有能力状态 S时,没有必要在此状态下继续发动一些不能获得更多能力的攻击,因为这样的攻击除了浪费精力和暴露自己,没有任何意义。至于一些类似 IP伪装的欺骗性攻击,在实验中,根据定义 1将另行构建入侵模型,这些欺骗性攻击最终都会被关联算法当成虚警排除。根据假设 1,那些不能给攻击者提供新的能力的警报信息往往是由于多点检测和网络延迟造成的误报和重复报警,在关联过程中直接滤去。此外,输入警报流是按照时间顺序对操作员汇报,因此一个新输入的超警报 h,其开始时间总是大于上一条警报信息的结束时间。

算法 1 警报关联算法

Step1 判断 h是否可以加入已存在的某个元攻击 M中。该元攻击的 C_set中元素的合取蕴涵 H_Consequence,若存在,则采用选择 3处理;若不存在,转 Step2。

Step2 判断 H_Prerequisite是否为空,若为空,用选择 1处理,若不为空则转 Step3。

Step3 判断元攻击 M中是否存在一个或多个超警报联合能构成 h的充分准备,即是否存在 M,满足 M的 C_set中元素的合取蕴涵 H_Prerequisite,若有,转选择 2,若没有,转 Step4。

Step4 元攻击合并算法。以 H_Prerequisite为搜索条件,判断是否存在多个元攻击,其 C_set的合取蕴涵 H_Prerequisite,若有,合并这些元攻击成为新的 M,转 Step5;若没有,用选择 1处理。

Step5 判断该新合并的元攻击 C_set中元素的合取是否蕴涵 H_Consequence,若是,采用选择 4处理,若不是,采用选择 2处理。

选择 1 创建新的元攻击 M,将 h加入 H_set,将 H_Consequence加入 C_set,确定时间戳;跳出,接受新输入的超警报,转 Step1。

选择 2 将 h加入已有元攻击 M,将 h放入相应 M的 H_set,同时将 H_Consequence放入 C_set,更改 M的时间戳;跳出,接受新的超警报,转 Step1。

选择 3 丢弃警报信息 h,基于假设 1,此为无效或者重复报警,直接丢弃;跳出,接受新输入的超

警报, 转 Step 1。

在 Step 4 中, 对于一些确实发生的警报, 虽然 IDS 提供警报信息显示不具备该攻击发生的充分准备条件, 只能认为是传感器或网络故障, 发生了漏检, 为此创建新的 M, 可忽略掉前面的漏检信息, 继续关联后续警报, 并构建攻击模型。

2.3 元攻击与入侵模型的一致性

由于元攻击本身就是对入侵模型的形式化描述, 因此, 元攻击与入侵模型之间是可以相互转化的, 从而证明了可以用算法 1 构建入侵模型; 此外, 由于这种转化是唯一的, 从而证明了元攻击与入侵模型的一致性问题。

元攻击聚集了输入警报流中的关联超警, 但是, 这种聚合只是简单的聚集, 没有显式表示这些超警报所代表的攻击之间的逻辑联系。

定义 7 如果 $M=\{H\_set, C\_set, time\}$ 是一个元攻击,

$$R=\{<h_1, h_2>|(h_1 Prep for h_2)\wedge (h_1, h_2\in H\_set)\} \quad (7)$$

则称 R 为元攻击 M 上准备关系。

定理 1 元攻击 M 上的准备关系是拟序关系。

证明 对于任意的元攻击 M, 假设 R 是定义在 M 上准备关系, 可证 R 满足:

1) 反自反性。对于任意的 $h_1\in M$, 显然, $A_1\_Time\_begin<A_1\_Time\_end$ 所以不满足 $A_1\_Time\_end<A_1\_Time\_begin$, 从而不可能有 $h_1, h_1\in R$, 其中 $A_1\_Time\_begin$ 和 $A_1\_Time\_end$ 分别是超警报 h_1 中组元 Aler 的开始时间和结束时间;

2) 反对称性。对于任意的 $h_1, h_2\in R$, 由定义 6 $A_1\_Time\_end<A_2\_Time\_begin$, 由于 $A_2\_Time\_begin<A_2\_Time\_end$, 所以不可能满足 $A_2\_Time\_end<A_1\_Time\_begin$, 从而不可能有 $h_2, h_1\in R$;

3) 可传递性。对于任意的 $h_1, h_2\in R$, $h_2, h_3\in R$, 由定义 6 $A_1\_Time\_end<A_2\_Time\_begin$ 且 $A_2\_Time\_end<A_3\_Time\_begin$, 从而 $A_1\_Time\_end<A_3\_Time\_begin$, 又由定义 6 $Pre_2\Rightarrow Cons_1$ 且 $Pre_3\Rightarrow Cons_2$, 由引理 1 可得 $Cons_2\Rightarrow Pre_2$, 所以 $Pre_3\Rightarrow Cons_1$, 从而根据定义 6 有 $h_1, h_3\in R$, 其中, Pre_i 和 $Cons_i$ 分别是超警报 h_i 的 Prerequisite 和 Consequence。

综合上述 3 点可得, M 上的准备关系 R 是拟序关系。

定义 8 R 是定义在元攻击 M 上的准备关系, $h_1, h_2\in R$, 且不存在 M 中的超警报 h_3 同时满足 $h_1, h_3\in R$ 和 $h_3, h_2\in R$, 则称 h_1 是 h_2 的直接准备。集合

$$COV(M)=\{h_1, h_2 | h_1, h_2\in R \text{ 且 } h_1 \text{ 是 } h_2 \text{ 的直接准备}\} \quad (8)$$

称为 M 上准备关系 R 的一个覆盖。

引理 2 对一个元攻击 M, 存在且仅存在一个关于准备关系的覆盖。

证明 从存在性和唯一性两方面来证明。

1) 对于任意一个元攻击 M, 按照定义可构建其上的准备关系 R, 和关于该准备关系 R 的覆盖 COV(M)。所以, 一个元攻击存在关于超警报准备关系的覆盖。

2) 假设有 2 个集合 COV_1 和 COV_2 都是元攻击 M 的覆盖, 然而对于任意的 $<h_1, h_2>\in COV_1$, 由 $<h_1, h_2>\in COV_1$ 知, $<h_1, h_2>\in R$ 且不存在 M 中的超警报 h_3, 同时满足 $<h_1, h_3>\in R$ 和 $<h_3, h_2>\in R$, 而由覆盖的构造方法知, 必然有 $<h_1, h_2>\in COV_2$, 同理, 对于任意的 $<h_1, h_2>\in COV_2$, 必然有 $<h_1, h_2>\in COV_1$, 因此 $COV_1=COV_2$。从而每个元攻击的覆盖是唯一的。

由此可知, 对于一个元攻击 M, 有且仅有一个与之对应的覆盖。

由引理 2 可知, 存在唯一的覆盖和元攻击对应, 可以将元攻击 M 转化成与它唯一对应的覆盖 COV(M), 然后通过 COV(M) 构建入侵模型。

算法 2 入侵模型构建算法

Step 1 任选一个属于 M 的超警报 h_i, 将 h_i 的准备条件 Pre 作为状态节点放在平面上, 将它的结束状态 Cons 节点放在 Pre_i 节点的右边, 并用从左到右的有向弧 h_i 连接 2 个状态节点。

Step 2 对于已存在的有向弧 h_i, 如果存在 $<h_i, h>\in COV(M)$, 则 h_i 的准备条件 Pre_i 作为状态节点放在 Pre 的左边, 画一条从 Pre_i 到 Pre 的有向弧; 如果存在 $<h, h_i>\in COV(M)$, 则 h_i 的结束能力 $Cons_i$ 作为状态节点放在 Cons 的右边, 画一条从 Cons 到 $Cons_i$ 的有向弧。

Step 3 重复进行 Step 2, 直到穷尽元攻击 M 中所有的超警报。

引理 3 对一个元攻击 M, 必存在一个基于状态转移的入侵模型, 且同一个元攻击的不同入侵模型是同构的。

证明

1) 存在性。对于给定的元攻击 M, 由引理 2可以构建与之唯一对应的关于准备关系的覆盖 COV(M), 采用算法 2, 可获得入侵状态转移模型。

2) 唯一性。对于同一个元攻击的多个入侵状态转移模型, 由于其节点集 C_set是相同的, 边集 H_set也是相同的; 节点与边的对应关系由 H_set中的元素 (超警报)所确定, 而相同的集合 H_set的元素必定也是完全相同的。所以, 这些属于同一个元攻击的多个入侵状态转移模型是同构。

定理 2 元攻击 M相应的覆盖 COV(M)和基于攻击者能力状态转移的入侵模型之间存在一一对应关系。

证明

1) 对于给定的元攻击 M, 由引理 2, 可找到唯一存在的 COV(M);

2) 由引理 3, 可以构建与之唯一对应的入侵状态转移模型;

3) 对于同一个元攻击的多个入侵状态转移模型, 初始化一个元攻击模型 M= { H_set, C_set, Time}, 把入侵模型中所涉及到的边 (超警报)放入 H_set, 状态节点所包含的能力等放入集合 C_set, 然后找到开始时间和结束时间确定时间戳 Time, 该 M唯一存在, 且和入侵模型一致。

综上所述, 可得到图 2描述的入侵模型和元攻击之间的转换关系, 元攻击 M通过 COV(M)可求得入侵模型, 而入侵模型可以形式化描述为元攻击, 因此这三者之间存在着一一对应关系。

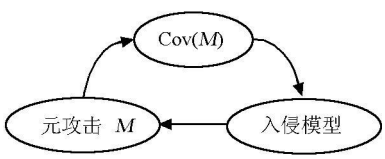

图 2 元攻击与入侵模型的转化关系

Fig. 2 Transform between meta-attack and intrusion model

3 混合式贝叶斯网络入侵预测方法

由于贝叶斯网络可根据观察到的片断、琐碎的现象推理出具有合理的因果联系, 完整而全面的描述整个事件的过程, 即构建良好的贝叶斯网络既能补充一些未观察到而又实际发生的现象, 还可以预测未来, 合理地推出主体未来可能采取的动作。这种能力用于入侵预测, 可以弥补入侵检测系统漏检造成的攻击行为信息丢失, 又可预测未来的攻击行为, 为容侵系统的快速主动响应提供判断依据。

3.1 混合式贝叶斯网络模型

对于图 1的入侵模型, 虽然在不同的参数情况以及攻击模式下, 即使采集到的警报信息相同, 攻击者采取的攻击行为以及该行为对系统的安全状态影响也有所不同。但是, 这里仍然蕴涵着一定的统计特征, 可以用贝叶斯网络进行特征统计与预测。

笔者提出一种新的混合式贝叶斯网络, 来描述定义 1中的入侵模型, 如图 3所示。网络结构分为安全状态和攻击行为两层, 定义 1中的状态节点置于安全状态层, 定义 1中所表示的攻击转化为攻击行为层节点。图 3中不同的层中节点之间的连接类型不同, 在层内, 同种节点间的连接为连续连接; 在层间不同节点间的连接为收敛连接。

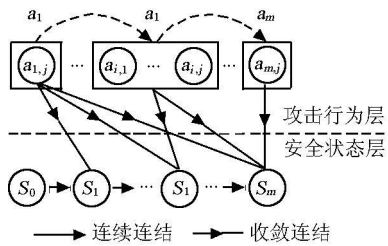

图 3 混合式贝叶斯网络

Fig. 3 Hybrid bayesian network

定义 9 混合式贝叶斯网络由节点和连接组成, 其节点分为状态节点和行为节点两类, 连接方式分为连续连接和收敛连接两类。

1) 状态节点 S_i表示系统的安全状态, 是攻击者能力的集合, 攻击者能力由定义 1给出。

2) 行为节点 a_i是攻击行为集合, 该集合中任一成员 a_i奏效后, 都会使得系统安全状态转移到 S_i或 S_{i+1}等后续安全状态。

3) 状态节点间是连续连接, 而行为节点到状态节点的连接为收敛连接。

根据图 3, 安全状态层的节点 S_i表示攻击者具备的对系统的操控能力, 即系统安全遭受的破坏; 安全状态之间的连接为连续连接, 如果不知道连接中变量的状态, 证据可以在连续连接中传送。已知 S_i发生, 会增加对 S_{i+1}的信念; 已知 S_{i+1}发生可确信状态 S_i已经发生。但是, 已知 S_{i+1}后, 再发生 S_i将不会增加对 S_{i+2}的信念, 即安全状态 S_i和 S_{i+2}是依赖的, 收到关于 S_{i+1}的证据时, 变成独立的。更形象的说, 安全状态层的有向连接反映了安全状态逐步下

降的过程, 当确信系统安全状态已处于 S_i时, 更加相信攻击者即将发动攻击, 使得自己具有 S_{i+1}以及其以后的状态所表示的能力; 然而, 当得知攻击者已经具有某能力后, 再知道它以前曾经拥有的攻击能力, 对预测未来的攻击是没有任何影响的。即系统安全状态与其先辈节点 (曾经经历过的安全状态) 相关, 而在已知其父节点 (前一个状态) 时, 则与其他先辈无关, 在原理上符合假设 1。

攻击行为层的节点 a_i表示攻击模型中的边集, 即来自入侵检测系统的表示攻击者行为的警报信号集合, 该集合中有 l_k个代表警报信号的节点 $a_{i,j}$($j=1, \cdots, l_k$), 当这些警报所代表的攻击行为奏效后, 系统安全状态立即下降至 S_i。从而, a_i中攻击奏效后, 系统安全状态将处于 S_i或者比 S_i更不安全的状态 S_k($k>i$)。攻击行为集合 a_i中任一节点与 a_{i+1}中任意节点之间都具有连续连接, 其意义同前面对安全状态节点中连续连接的描述。a_i中任意 2 个节点 $a_{i,j}$和 $a_{i,k}$($j, k=1, \cdots, l_k$)与它可获得的下一个状态 S_i之间的跨层连接是收敛连接。即如果关于 S_i的状态一无所知, 那么 S_i的任意 2个双亲节点 (可以获得 S_i的任意 2 个攻击行为 $a_{i,j}$和 $a_{i,k}$) 是独立的, 即 $a_{i,j}$的知识对 $a_{i,k}$没有影响, $a_{i,k}$的知识对 $a_{i,j}$也没有影响。但是, 如果 S_i的状态已知了, 那么 $a_{i,j}$和 $a_{i,k}$就变成了条件依赖的。对其中一个攻击的信念上升时, 对另一个的信念就会下降, 这同样和假设 1相符。

3.2 联合概率分布计算

如果给定父节点, 可以计算出每个子节点的条件概率分布。如果被观察节点没有父节点, 那么就使用先验值作为条件概率分布。图 1中如果 $i>j$ 状态 S_i时攻击者所拥有的能力会包含 S_j状态下攻击者所拥有的能力, 但是, 考虑到当攻击者拥有相应能力, 进入状态 S_i后, 不能再认为处于状态 S_j, 就像一个人经过童年进入中年后, 不能再认为他处于童年状态一样, 为此, 假设这些状态 $\{S_0, S_1, \cdots, S_m\}$ 构成系统安全状态空间 Ω的理想完备划分, 即

$$\sum_{i=1}^{m} S_i = \Omega \text{ 且 } S_i \cap S_j = \varphi, \quad i \neq j \tag{9}$$

定义以下变量:

1) 变量的状态集合 $S=\{S_0, S_1, \cdots, S_m\}$, S_0表示初始安全状态, S_i表示在目的变量的状态集合中第 i个系统安全状态;

2) 者为达到攻击能力 S_k而采取的攻击步骤 a_k奏效后, 攻击者将拥有攻击能力 S_i, 即系统的安全状态下降为 S_i;

3) $k=\{S_0, a_1, a_2, \cdots, a_k\}$表示系统从初始状态 S_0开始发动的攻击 $a_1, a_2, \cdots, a_k$一段长度为 k的攻击序列;

4) $i-1, j, l_k$($k=1, 2, \cdots, l_k$) 表示在长度为 m的一条攻击序列中, 在第 $i-1$个安全状态下, 攻击者转而采取攻击行为 $a_{i,j}$的转移概率;

5) 表示事件空间, 即完备的攻击行为的事件空间。

根据图 3中混合式贝叶斯网络, 关于状态与攻击的信念, 可以得到如下的形式化的基本规则:

1) 达到同一状态的所有攻击发生的信念和记为

$$P(a_i) = \sum_k P_{i-1, j, k} \tag{10}$$

2) 对某一状态 S_i能进入该状态的所有攻击发生的信念和为 1, 即

$$\sum_k (P_{i-1, j, k} \mid S_i) = 1 \tag{11}$$

3) 当观察到 $a_{i,k}$时, 进入 S_i状态的概率增加;

4) 当系统安全状态在 S_{i-1}时, 信念值 $P(a_i)$和 $P(S_i)$要增大。

对于某个从初始状态 S_0开始的攻击序列 $A_k=\{a_1, a_2, \cdots, a_k\}$中, 由于攻击 a_i的成功, 意味着系统进入状态 S_i记为

$$\sum_{i=1}^{m} (A_k \cap S_i) = A_k \cap \Omega = A_k \tag{12}$$

由贝叶斯网络中蕴涵的马尔可夫假设, 可得

$$(a_k \mid S_i, a_1, a_2, \cdots, a_{k-1}) = (a_k \mid S_i, a_{k-1}) \tag{13}$$

于是有

$$A_k \cap S_i = \{S_0, a_1, a_2, \cdots, a_k\} = \{S_0, a_1, a_2, \cdots, a_{k-1}\} \cap (a_k \mid S_i, a_1, a_2, \cdots, a_{k-1}) = (S_i \mid A_{k-1}) \cap (a_k \mid S_i, a_{k-1}) \cap A_{k-1} \tag{14}$$

这里 $(S_i \mid A_{k-1})$表示 S_i是 A_{k-1}的目的或攻击者希望获得的能力, 或系统安全将下降到的状态; $(a_k \mid S_i, a_{k-1})$表示在攻击者使用 a_{k-1}处于状态 S_{i-1}后, 下一步采取攻击行为 a_k进入 S_i的概率。此时, 由全概率公式得

$$P(A_k) = P\left(\sum_{i=1}^{m} (A_k \cap S_i)\right) = \sum_{i=1}^{m} (A_k \cap S_i) \tag{15}$$

再由条件概率公式得

$$P(S_i \mid A_k) = P(A_k \cap S_i) / P(A_k) = P(S_i \mid A_{k-1}) P(a_k \mid S_i, a_{k-1}) P(A_{k-1}) / \sum_{i=1}^{m} P(S_i \mid A_{k-1}) P(a_k \mid S_i, a_{k-1}) P(A_{k-1}) = P(S_i \mid A_{k-1}) P(a_k \mid S_i, a_{k-1}) / \sum_{i=1}^{m} P(S_i \mid A_{k-1}) P(a_k \mid S_i, a_{k-1}) \quad (16)$$

这就得到了贝叶斯后验概率公式, 保证了基于贝叶斯网络推理的正确性。

4 实验描述

为验证模型及预测算法的有效性, 使用 MIT Lincoln Lab提供的 DARPA2000入侵检测攻击场景数据集 LLDOS1.0, 测试所提出的入侵模型以及基于该模型的入侵预测算法的性能。 LLDOS1.0中包含了一个完整的攻击序列, 主机探测、端口扫描、系统入侵、安装木马, 利用被控主机发起 DDOS攻击。在测试中, 每一个数据集包在 3 h内采集到评测网络的 DMZ和内部网的数据流量, 背景流量来自 DARPA99给出的不含攻击的训练数据, 通过 Netpoke软件实施攻击重放, 将这些流量导向 RealSecure6.0等网络 IDS, 利用这些 IDS充当传感器, 将警报数据传向容侵系统的触发系统。

为达到系统入侵的目的, 通常攻击者采取 5个步骤: 地址扫描、端口扫描、口令文件获取、口令破解和登陆系统, 对于系统中某主机而言, 其安全状态逐渐下降, 依次为初始安全状态 (S_0)、地址被探测 (S_1)、端口及相应服务被探测 (S_2)、口令文件被获取 (S_3)、口令被破解 (S_4)以及系统被控制 (S_5), 其入侵过程的混合式贝叶斯网络如图 4所示。

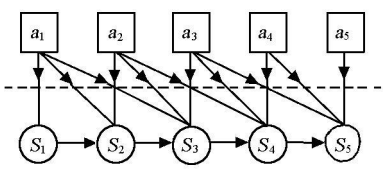

图 4 系统入侵过程的混合式贝叶斯网络表示

Fig. 4 HYBN for intrusion process

在网络推理过程中, 为保证

$$\sum_{i=0}^{m} P(S_i \mid A_k) = 1 \quad (17)$$

对处于初始状态的系统, 设置信度为 $P(S_0) = 1$, $P(S_i) = 0$ ($i = 1, \cdots, 5$), 如果发现集合 a_{l+1}中警报 a_l使得攻击序列 A_{l-1}成为 A_l, 表明攻击者已实施相关攻击, 则系统安全状态已下降至 S_l, 也有可能有攻击行为未被发现或即将实施, 而系统状态已经或将处于 S_{l+1}或 S_{l+2}。为此, 当 $k < m-1$时, 采用

$$P(S_k \mid A_l) = \begin{cases} 0 & k < l \\ (k-1)/k & k = l \\ 1 - \sum_{i=0}^{k-1} P(S_i \mid A_l) \text{和 } 1/k \text{中的较小者,} & k > l \end{cases} \quad (18)$$

在 $l = m-1$时, 采用

$$P(S_k \mid A_l) = \begin{cases} 0 & k < l \\ (k-1)/k & k = l \\ 1 - \sum_{i=0}^{k-1} P(S_i \mid A_l), & k > l \end{cases} \quad (19)$$

进行信度更新。

在具体的实验中, 信度由 $P(S_0) = 1$开始, 当攻击导致安全状态变化时, 采用式 (18)、式 (19)进行信度更新。依次为 $P(S_1 \mid A_1) = 1/2$, $P(S_2 \mid A_1) = 1/3$, $P(S_3 \mid A_1) = 1/6$且记 $P(S_0 \mid A_1) = 0$表示系统已不可能回到最初的安全状态; 此时, 若又有警报表明 a_2中攻击已奏效, 则第二次转变为 $P(S_0 \mid A_2) = 0$, $P(S_1 \mid A_2) = 0$, $P(S_2 \mid A_2) = 3/4$, $P(S_3 \mid A_2) = 1/5$, $P(S_4 \mid A_2) = 1/20$。这样赋值, 有一定的事实依据, 首先 l值越大, $P(S_k \mid A_l)$的初值越大, 攻击者已完成一个攻击序列中前面的准备环节越多, 且正在发动下一个环节的攻击, 成功的概率会更大; 同样, 随着 l值增大, 攻击的危险级别越高, 但对攻击者发动此类攻击的难度也在增大, 因而 $P(S_k \mid A_l)$ ($k > l$)越来越小。当有少量的警报发生时, 可根据贝叶斯网络中的连接方式进行信度调整。

在应用中, 根据系统要求的安全级别来设定阈值, 通过判断系统安全状态的信度是否超过该阈值来决定是否需要采用响应措施。实验中分别在 DMZ和内部网释放了 89和 60次攻击, RealSecure分别产生了 891和 922个警报, 这些警报包含了大量的虚警和重复报警, 根据最大概率是否超过阈值来预测该系统的下一步安全状态, 并将实时预测结果和正确报警所确定的真实状态相比较, 验证算法的可行性。实验过程中发现 0.81是一个较好的阈值, 此时得到的预测结果较为理想。

由于尚不具有统一的系统平台、测试数据集以及定义在其上的类似于预测率的衡量指标, 一些相关的研究工作都未曾给出可比较的结果。

Cheung[9]和 Ning[11]的警报关联方法需要 2～180 s的离线处理，才能发现被漏报的信息；而基于混合式贝叶斯网络的入侵预测方法可以在线预测后一个攻击行为对系统安全可能造成的影响，所得到的预警信息均在 RealSecure警报来临之前，能达到提前预测的功能。由于目的是给容侵系统提供更多的响应时间，从这个意义上看，基于混合式贝叶斯网络的入侵预测方法，具有更好的效果。

5 相关工作比较

与本文相关的主要工作有警报关联[8～10]、入侵预测[11～13]。文献[8～10]中的警报关联工作与入侵模型构建工作相类似，文献[8～10]重在关注攻击行为的内部关联，而 HYBN入侵模型是为反应式容侵系统提供预测和决策依据，关注攻击行为对系统造成的影响。为此 HYBN的算法构建的是基于攻击者能力的入侵模型，通过状态变迁反映系统安全状态的下降过程。此外，给出了相关证明，从理论上保障了该模型构建方法的可行性。

关于入侵预测的工作，Ramasubramanian[12]旨在讨论针对数据库的异常行为，通过统计发现入侵行为的规律，从而提供预测依据，从本质上说，这些工作依然是基于异常行为的入侵检测方法。Qin[13]对攻击树进行统计分析，找出攻击规划模型，并通过规划模型使用因果网络进行警报预测；Wang[14]使用队列图进行警报关联，然后根据队列图，预测攻击者的下一步攻击行为。由于尚不具有统一的系统平台、测试数据集，从而无法从实验的角度提供 HYBN方法与这两类方法的比较。

举例说明如下，如图 4所示的系统入侵过程中，如果地址探测、端口扫描、获取口令文件、口令破解和登陆系统 5个步骤中，攻击者为使得系统安全状态转移到 S_f（$f=1, \cdots, 5$），以此可以有 3，4，2，3，1种攻击方式可供选择，如图 5所示，则攻击者可选择使用 3×4×2×3个攻击场景中的任意一个场景，Qin[13]的攻击树匹配方法需要对这 72个入侵场景进行逐一匹配，Wang[14]的方法则如图 6所示，分为 5个阶段进行预测，整个过程中有 3+4+2+3=12种选择，相对于 Qin[13]的方法具有了很大的改进。而对 HYBN方法来说，只需要判断处于 5个安全状态中的哪一个即可，如图 4所示，推理的结果更为简单，由于安全状态迁移在时间上的不可逆性，复杂度为 4+3+2+1=10。

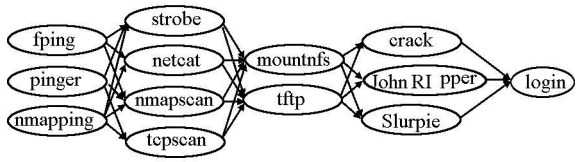

图 5 入侵过程的场景关联图

Fig. 5 Intrusion scenario by correlating alerts

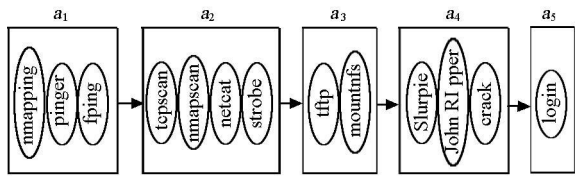

图 6 入侵过程的队列图

Fig. 6 Queue for intrusion process

定理 3 设入侵过程中系统共有 m个安全状态，分别为 S_1，S_2，…，S_m，可获得状态 S_i的攻击行为集合为 a_i，则该入侵过程可以选择 $\prod_{i=1}^{m}|a_i|$个攻击场景，队列图方法中共有 $\sum_{i=1}^{m}|a_i|$种可能的预测结果，而混合式贝叶斯网络中共有 $\sum_{i=1}^{m-1}$ 可能的预测结果。其中 $|a_i|$是集合 a_i的模，表示其中攻击行为的数目。

为了讨论和分析的方便，在定理 3中假设集合 a_i中都有相同数量的攻击，则可以得出如下推论：

推论 1 设基于攻击者能力的入侵模型中共有 m个中间状态，分别为 S_1，S_2，…，S_m，可获得状态 S_i的攻击行为集合为 a_i，又设 $|a_i|=n$，即集合 a_i中都有相同数量的攻击，则共有 n^m 个可能的攻击场景，而使用队列图需要 mn次预测，使用混合式贝叶斯网络需要 $\sum_{i=1}^{m-1}$ 次预测。

假设入侵过程中有 5个攻击步骤，即 $m=5$，在具体的情况下，每个步骤可以有多种攻击方式，在实际情形中，还可以通过变换源地址而形成新的警报信号，从而攻击手段 n可以有多种选择，通常会较大。图 7给出了 3种算法随 n从 1增大到 6时匹配空间增长的情况，通过对比，很容易看出，在 m固定时，Qin[13]的攻击树匹配方法匹配的次数与攻击场景数相等，随着 n的增长而呈 $O(n^5)$增长；Wang[14]的队列图匹配方法随 n呈线形增长；而 HYBN方法不因为 n的增长而变化。同理，由推论 1可知，若固

定 n,则 Qin[13]的攻击树匹配方法随着 m的增长而呈指数增长; Wang[14]的队列图匹配方法和 HyBN方法随 m呈线形增长。这和第 4节实验提供的数据相吻合。

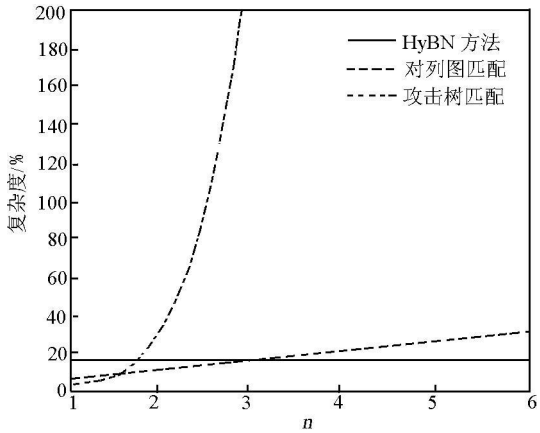

图 7 m = 5时 3种方法的复杂度比较

Fig. 7 Complexities of three methods with m = 5

6 结语

为给反应式容侵系统提供响应依据,研究适用于容侵系统的入侵预测方法。采用基于攻击者能力的入侵模型,将入侵描述为攻击者能力不断提高、系统安全状态不断下降的过程,有利于为容侵系统提供触发引擎;提出了构建该入侵模型的方法,并给出了该构建方法是可行性的理论证明。在入侵模型的基础上,提出了一种混合式贝叶斯网络(HyBN)模型及相应的联合概率分布计算方法;在相关信度更新算法的支持下,实验证实了所提出算法的可行性,分析表明了该算法具有大大低于同类工作的时间复杂度。

致谢 感谢博士后合作导师顾冠群院士、罗军舟教授及实验室研究生在论文完成阶段的帮助,感谢解放军信息工程大学电子技术学院郭渊博博士在入侵模型方面提供的启发。

参考文献

[1] Verissimo P E, Neves N F, Correia M P. Intrusion-tolerant Architectures: Concepts and Design[J]. Lecture Notes in Computer Science, 2003, 2677: 90-109

[2] 崔竞松, 王丽娜, 张焕国, 等. 一种并行容侵系统研究模型——RC模型[J]. 计算机学报, 2004, 27(4): 500-506

[3] Marsh A, Schneider B. CODEX: a robust and secure secret distribution System[J]. IEEE Transactions on Dependable and Secure Computing, 2004, 1(1): 34-47

[4] Kursawe K. Asynchronous Byzantine Group Communication[C]. 21st IEEE Symposium on Reliable Distributed Systems, Osaka, Japan, 2002: 352-357

[5] Liu Peng, Jing Jiwu. The Design and Implementation of a Self-Healing Database System[J]. Journal of Intelligent Information Systems, 2004, 23(3): 247-269

[6] Wang Rong, Wang Feiyi, Gregory B. Design and implementation of acceptance monitor for building intrusion tolerant systems[J]. Software-Practice and Experience, 2003, 33(14): 1399-1417

[7] Julisch K. Mining alarm clusters to improve alarm handling efficiency[A]. 17th Annual Computer Security Applications Conference (ACSAC'01)[C]. New York, 2001: 12-21

[8] Undercoffer J, Pinkston J. Modeling computer attacks: a target-centric ontology for intrusion detection[A]. The Sixth International Symposium on Recent Advances in Intrusion Detection[C]. Pittsburgh, PA, USA, 2003

[9] Cheung S, Lindqvist U, Martin W. Modeling multi-step cyber attacks for scenario recognition[A]. DARPA Information Survivability Conference and Exposition[C]. Washington, DC, 2003

[10] Zhou Jimmy, Heckman M, Reynolds B, et al. Modeling network intrusion detection alerts for correlation[J]. ACM Transactions on Information and System Security, 2007, 10(1): 1-31

[11] Ning Peng, Cui Yun, Reeves D S, et al. Techniques and tools for analyzing intrusion alerts[J]. ACM Transactions on Information and System Security, 2004, 7(2): 274-318

[12] Ramasubramanian P, Kannan A. Quick-prop neural network short-term forecasting framework for a database intrusion prediction system[J]. Lecture Notes in Artificial Interlligence, 2004, 3070: 847-852

[13] Qin Xinzhou, Li Wenke. Attack plan recognition and prediction using causal networks[A]. Proceeding of 20th Annual Security Application Conference (ACSAC'04)[C]. Tucson, Arizona, December 2004

[14] Wang Lingyu, Liu Anyi, Sushil J. An efficient and unified approach to correlating, hypothesizing, and predicting intrusion alerts[A]. Proceedings of the 10th European Symposium on Research in Computer Security (ESORICS2005)[C]. Germany: Springer Press, 2005: 247-266

Hybrid bayesian network method for predicting intrusion in reactive intrusion tolerance system

Wang Liangmin[1,2], Ma Jianfeng[3]

(1. School of Computer Science and Engineering of Southeast University, Nanjing 210018, China; 2. School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China; 3. Key Lab of Computer Network and Information Security of Education Ministry, Xidian University, Xi'an 710071, China)

[Abstract] To solve the open problem of predicting intrusion in Reactive Intrusion Tolerance System, a hybrid Bayesian network method is presented in this paper. Firstly, an intrusion model is presented, which pays its emphasis on the influence of the intrusion upon the system and describes the intrusion as the state transition process of the attackers' capability. The intrusion model is appropriate to trig the reactive intrusion tolerance system. We proposed the constructing algorithm and the proof of its feasibility. Secondly, a hybrid Bayesian network model based on this intrusion model is presented to show the casual relation of the attack behavior and secure state. The model is divided into two layers: attack behavior layer and secure state layer, in which the attack edges and state nodes of intrusion model are used as nodes in behavior layer and state layer respectively. In this hybrid Bayesian network model, the connections of the same layer are continuous, but that of the different layer are converge. The algorithm for computing the joint probability distribution of the hybrid Bayesian network is presented. In the end, the efficiency of the intrusion model and hybrid Bayesian network in predicting intrusion is shown by the experiment with our belief propagation algorithm, and the advantages of this predicting method over the related work are shown by analysis and comparisons.

[Key words] intrusion tolerance; alert correlation; intrusion model; intrusion prediction