考虑时间参数的网络安全态势评估模型研究*

孟 锦[1], 徐 佳[2], 何加浪[1], 张 宏[1]
(1. 南京理工大学 计算机科学与技术学院, 南京 210094; 2. 南京邮电大学 计算机科学与技术系, 南京 210003)

摘 要: 针对已有网络安全态势评估由于缺乏考虑数据源时变性而造成的证据源证据不可靠、评估出现误差的问题,引入时变函数来刻画多源证据的时变性,提出时变 D-S 证据理论,用以提高证据的可靠性。使用改进的时变 D-S 证据理论方法对多传感器的证据进行融合,得到威胁的发生概率,并在此基础上提出带有时间参数的网络安全态势评估层次化模型;最后利用网络实例数据,对所提出的网络安全态势评估模型进行了验证。实验对比结果表明,该模型对实际网络运行情况的评估更符合实际情况。

关键词: 网络安全; 态势评估; 时变 D-S 证据理论; 多传感器数据融合; 评估模型
中图分类号: TP393.08 **文献标志码**: A **文章编号**: 1001-3695(2012)10-3820-04
doi:10.3969/j.issn.1001-3695.2012.10.057

Network security situation assessment model based on time parameter

MENG Jin[1], XU Jia[2], HE Jia-lang[1], ZHANG Hong[1]
(1. *School of Computer Science & Technology, Nanjing University of Science & Technology, Nanjing* 210094, *China*; 2. *Dept. of Computer Science & Technology, Nanjing University of Posts & Telecommunication, Nanjing* 210003, *China*)

Abstract: For resolving the problem that due to the lack of consideration of time-varying of data sources, the existing network security situation assessment usually failed to make the source of evidence reliable enough and error existed in the assessment results, this paper introduced the time-varying function to describe the time-varying multi-source evidence. It proposed the time-varying D-S evidence theory to improve the reliability of the evidence. The occurrence probability of the threat was obtained by using the improved time-varying D-S evidence theory to multi-sensor fusion of the evidence. On this basis, it proposed the hierarchical model with time parameter. Finally it gave an example of the actual network datasets to valiadate the network security situation awareness model. The results of the experimental comparison show that this model is more in line with the actual situation.
Key words: network security; situation assessment; time-varying D-S evidence theory; multi-sensor data fusion; evaluation model

0 引言

网络安全态势感知(network security situational awareness)是目前网络安全领域的一个研究热点,已经引起了相关科研机构和研究人员的足够重视。随着网络规模的不断壮大, 网络结构日益复杂, 网络病毒、DoS/DDoS 攻击等构成的威胁和损失越来越大, 传统的、单一的防御设备或者检测设备已经无法满足安全需求, 因此迫切需要新的技术来对网络安全状况进行实时监控和预警。网络安全态势评估技术能够综合各方面的安全因素, 融合防火墙、防病毒软件、入侵检测系统(IDS)、安全审计系统等安全措施的数据信息,从整体上反映网络安全状况, 对当前的网络安全状态进行定量和定性的评估,为增强网络安全性提供可靠的参照依据。国内外的研究人员依据不同的技术思路,设计并实现了大量针对计算机网络的安全态势评估方法。

近年来,多异质传感器环境中实现网络安全态势感知的研究也已受到重视,目前已有的研究[1~6]为网络安全态势评估模型及算法的研究奠定了良好的基础,但同时也普遍存在着一些缺陷,例如通常只考虑外部攻击信息,缺乏对主机内部信息的综合考虑;忽略了数据源证据的时变性,将证据作为全效证据使用,使得评估结果不够准确。针对这些问题,结合已有的研究成果,提出了基于时变 D-S 证据理论的网络安全态势评估模型,引入了时变函数来刻画多源证据的时变性,使用改进后的时变 D-S 证据理论方法对多传感器的证据进行融合,得到威胁的发生概率;提出带有时间参数的层次化模型,融合计算网络安全态势值,从而实现了网络安全态势的量化评估。通过实例对比可以明显看出,采用时变 D-S 证据理论对网络安全态势进行评估的模型对不考虑时间参数的评估模型的改进,评估结果更符合实验网络情况,网络安全态势评估的连续性和有效性得到了显著提高。

1 网络安全态势评估模型

1.1 模型基本定义

本文对于影响网络安全的各种攻击和安全问题,统称为威胁。

定义 1 威胁的发生概率。威胁的发生概率是将多个传感器收集的数据进行数据融合得到的威胁已经发生的可能性,表示为

收稿日期: 2012-04-16; **修回日期**: 2012-05-24 **基金项目**: 国家自然科学基金资助项目(90718021,60903027); 自主科研先期投入计划资助项目(2010XQTR04)

作者简介: 孟锦(1984-),女,江苏泰州人,博士研究生,主要研究方向为信息安全、风险评估、态势感知(mengjin1107@yahoo.com.cn);徐佳(1980-),男,讲师,博士,主要研究方向为网络安全、网络协议;何加浪(1984-),男,博士研究生,主要研究方向为可信软件;张宏(1956-),男,博导,博士,主要研究方向为信息安全、数据挖掘.

$M(h)$。其中:h 表示正确报警,即威胁发生;$\bar{h}$ 表示错误报警,即威胁不发生。

定义 2　威胁的严重度 R。威胁的严重度是威胁可能带来的后果的量化。采用 Snort 用户手册[7]提供威胁的分类和优先级划分方法,根据威胁类别把威胁严重度划分为 high、medium、low 三个等级,分别用 3、2、1 表示。

1.2　基于多源传感器证据融合的层次化网络安全态势模型

在评估模型方面,陈秀真等人[8]提出的层次化风险评估模型是基于 IDS 检测报警日志的,并结合服务、主机本身的重要性及网络的拓扑结构,提出采用自下而上、先局部后整体评估策略的层次化网络安全威胁态势量化评估方法。但其方法只考虑了检测设备 IDS 这一种报警日志来源,实际情况中,网络中部署着各类检测设备,如防火墙、NIDS、HIDS 等。参考文献[8]模型的层次化概念,加入多源传感器证据融合技术,提出基于多源传感器数据融合的层次化网络安全态势模型,如图 1 所示,采用自下而上、先局部后整体评估策略。

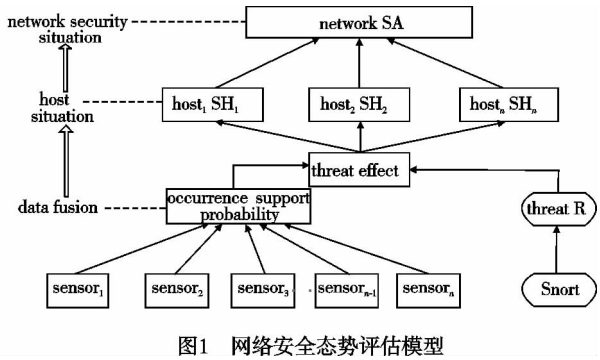

图1　网络安全态势评估模型

模型流程如下:

a)通过多传感器数据融合,得到威胁的发生概率 M,具体求解下文给出。

b)单个威胁对主机的影响度计算为

$$\mathrm{Sh} = M \times R \tag{1}$$

c)主机的安全态势值计算。主机的安全态势值量化为主机某一时间所受的所有威胁的影响力等于威胁的发生概率乘以威胁的严重程度。

$$\mathrm{SH} = \sum_{i=1}^{n} (M_i \times R_i) \tag{2}$$

其中,i 为主机所受的所有威胁数。

d)网络安全态势值 SA 计算。网络系统级的安全态势值,量化为网络系统某一时间所有主机所受的威胁的影响力。网络系统中各个主机所处的地位不同,也就决定了其在网络安全态势中所占的权重不同。因此

$$\mathrm{SA} = \sum_{i=1}^{n} w_i \mathrm{SH}_i \tag{3}$$

其中:i 为网络中所有的主机数,w_i 为每个主机的权重。

主机权重,即主机在网络中的地位 w,取决于主机提供的服务的重要性,主机提供重要程度越高的服务数越多,那么主机的地位越高,即主机权重越大。服务重要程度划分为高、中、低三个等级,网络系统各种服务的权重和为 1。则主机权重为

$$w = f_h N_h + f_m N_m + f_l N_l \tag{4}$$

其中:f_h f_m f_l 为服务的高、中、低三个重要程度对应的量化分值;N_h、N_m、N_l 为主机中三种等级服务所占的个数。

2　基于时变 D-S 证据理论的网络安全态势评估模型

多传感器数据融合是将网络中设置的安全检测系统所采集的数据进行融合的方法。由于不同的传感器即使在检测相同的攻击时也有着不同程度的精确度,因此,对于分散式的安全检测系统采集的数据不应该一视同仁。基于此,本文定义传感器对威胁的检测精度为威胁发生支持概率。采用 D-S 证据理论的合成规则来融合不同感应器的检测数据。

2.1　D-S 证据理论

D-S 理论最初是基于 Dempster 用概率范围去表示实际问题中的不确定性所做的工作,之后经过 Gshafer 的进一步拓展,形成了一种不精确推理理论,即 D-S 证据理论,可以用于融合多个证据源提供的证据[7]。

定义 3　设 Θ 为识别框架,如果集函数 m: $2^{\Theta} \to [0,1]$(2^{Θ} 为 Θ 的幂集)满足:a)$m(\varphi) = 0$;b)$\sum_{A \subseteq \Theta} m(A) = 1$,则称 m 为框架 Θ 上的基本概率分配函数。

定义 4　信度函数(belief function)Bel:$2^{\Theta} \to [0,1]$,对任意的 $A \subseteq \Theta$,有 $\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B)$。

定义 5　Demspter 合成公式

$$m(\varnothing) = 0$$

$$m(A) = \frac{1}{1-k} \sum_{\substack{A_1,\cdots,A_n \subseteq \Theta \\ A_1 \cap \cdots \cap A_n = A}} \prod_{1 \leqslant i \leqslant n} m_i(A_i), A \neq \varnothing \tag{5}$$

其中:$k = \sum_{\substack{A_1,\cdots,A_n \subseteq \Theta \\ A_1 \cap \cdots \cap A_n = A}} \prod_{1 \leqslant i \leqslant n} m_i(A_i)$,$k$ 的大小反映了证据冲突的程度,$\frac{1}{1-k}$称为归一化因子,作用是为了避免在合成时将非零的概率赋给空集。

由于 D-S 证据理论合成公式存在不足,在处理冲突证据时会出现违背常理的结论,所以采用文献[9]中改进的 D-S 证据理论方法对多传感器的 m 函数值进行合成。通过推导,可以得到 n 个传感器的合成公式为

$$\begin{aligned} m(\varnothing) &= 0 \\ m(h) &= \prod_{i=1}^{n} m_i(h) + k \times q(h) \\ m(\bar{h}) &= 1 - m(h) \\ m(H) &= 0 \end{aligned} \tag{6}$$

其中:$k = 1 - \prod_{1 \leqslant i \leqslant n} m_i(h) - \prod_{1 \leqslant i \leqslant n} m_i(\bar{h})$, $q(h) = \frac{1}{n} \sum_{i=1}^{n} m_i(h)$。

2.2　时变 D-S 证据理论

多传感器融合是对多个传感器采集的威胁相关数据进行融合,得到威胁的发生概率。不同的传感器其本身对于威胁具有不同的检测精度,这也就决定了其对于威胁发生概率的不同支持力度。

根据 D-S 证据理论定义,辨识框架 $\Theta = \{h, \bar{h}\}$,则幂集 $2^{\Theta} = \{\Phi, h, \bar{h}, H\}$,其中:$\Phi$ 表示威胁既发生,又没有发生;h 表示威胁发生;$\bar{h}$ 表示威胁没有发生;H 表示威胁可能发生,也可能没有发生。则传感器 i 的基本概率分配函数可表示为:$m_i(\varnothing) = 0$;$m_i(h)$;$m_i(\bar{h})$;$m_i(H) = 0$。

证据理论中,证据是经验和知识的一部分,是对问题所作的观察和研究的结果,其本身就是一个时变过程。各传感器对威胁的发生支持率也不可能保持不变,会随着时间的推移而变化。在对多证据源提供的证据进行分析时还发现,各个独立的

证据源间存在时间差,在以前的使用 D-S 证据理论进行网络安全态势评估中,并没有考虑证据信息的时变性问题,而是将所有的证据信息都校准到同一个时刻点 t,且所有的证据都是全效证据。本文提出用带时间参数的函数来描述基本概率分配值的变化规律,使得改进后的 m 值由静态值变为随时间衰减函数变化的动态值,从理论上讲更具备合理性和通用性,也更符合实际情况。

根据已有研究假设证据可信度按指数衰减。得到改进后的基本概率分配函数 m 为

$$m(\varnothing)=m(H)=0$$
$$m(h,t)=m(h,t_0)\mathrm{e}^{-\frac{t-t_0}{\Delta T}}\quad m(\bar{h},t)=1-m(h,t)\tag{7}$$

其中:t_0 是证据的生成时间,每个证据的生成时间是相互独立的;ΔT 是可调参数,用来改变证据随时间衰减的快慢程度。

证据合成规则推导。设 m_1 和 m_2 是 Θ 上的两个证据信息的带时间参量基本概率分配函数,时刻 t_1 和 t_2 分别为证据生成时间,那么在 t 时刻的证据合成计算方法为

$$m(\varnothing)=0$$
$$m(h,t)=m_1(h,t)\oplus m_2(h,t)=$$
$$m_1(h,t)m_2(h,t)+kq(h)=$$
$$m_1(h)\mathrm{e}^{-\frac{t-t_1}{\Delta T}}m_2(h)\mathrm{e}^{-\frac{t-t_2}{\Delta T}}+kq(h,t)$$
$$m(\bar{h},t)=1-m(h,t)$$
$$m(H)=0\tag{8}$$

其中:$k=1-m_1(h)m_2(h)-m_1(\bar{h})m_2(\bar{h})$,$q(h)=\frac{1}{2}(m_1(h)\mathrm{e}^{-\frac{t-t_1}{\Delta T}}+m_2(h)\mathrm{e}^{-\frac{t-t_2}{\Delta T}})$。

推广到 n 个证据源:

$$m(\varnothing)=m(H)=0$$
$$m(h,t)=\prod_{i=1}^{n}m_i(h,t)+kq(h,t)$$
$$m(\bar{h},t)=1-m(h,t)\tag{9}$$

其中:$k=1-\prod_{1\leqslant i\leqslant n}m_i(h)-\prod_{1\leqslant i\leqslant n}m_i(\bar{h})$,$q(h,t)=\frac{1}{n}\sum_{i=1}^{n}m_i(h,t)$。

通过上述推导计算,得到了某个威胁的发生概率 $m(h,t)$,依次代入式(1)~(3),即可求得网络安全态势值 SA。

3 实例分析

为了验证本文模型的有效性,本文使用 2000 年 DARPA 评估数据集作为实验数据。该数据集中提供了一个攻击场景 LLDOS 1.0,包含了五个步骤的攻击,本文针对此攻击场景进行基于时变证据理论的网络安全态势分析。DARPA 没有给出明确的网络拓扑图,根据对数据集的分析,构造出了如图 2 所示的网络图。

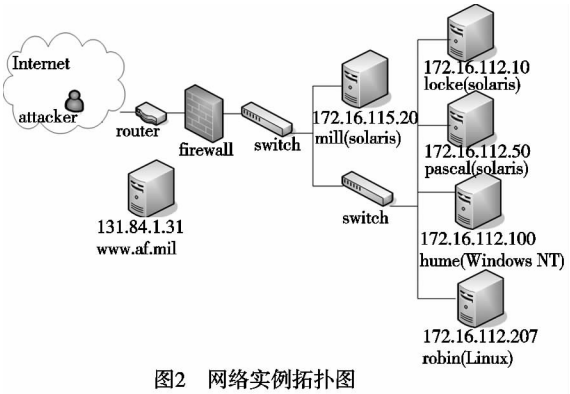

图2 网络实例拓扑图

结合主机信息、数据集报警信息和审计日志,将数据集 LLDOS 1.0 总共五个攻击步骤分为五个时段,利用上文态势评估算法,根据模型流程四个步骤,分别计算各个时间段的安全态势评估值。实验数据及结果如表 1 ~3 所示

表 1 网络实例中利用到的脆弱性列表

vulnerability name	mill	locke	pascal	hume	robin	www. af. mil
Sadmind buffer overflow	y	y	y	n	n	n
ICMP incorrectly cofigured	y	y	y	y	y	n
SYN Flood	n	n	n	n	n	y
RCP incorrectly cofigured	y	y	y	n	n	n
HINFO query incorrectly cofigured	y	n	n	n	n	n
SunRPC incorrectly cofigured	y	y	y	n	n	n

表 2 主机服务信息列表

service	weight(w)	mill	locke	pascal	hume	robin	www. af. mil
POP3	0.1	n	n	n	y	n	n
DNS	0.2	y	n	n	n	n	n
HTTP	0.2	n	y	y	n	y	y
FTP	0.2	y	y	y	n	n	n
SMTP	0.1	n	n	n	y	n	n
Telnet	0.2	y	y	y	n	n	n

表 3 实验计算结果

time	mill	locke	pascal	hume	robin	www. af. mil	SA
WH	0.34	0.18	0.18	0.2	0.05	0.05	1
t_0	0.036 7	0.04	0.036 7	0.04	0	0	0.034 3
t_1	0.047 3	0.051 5	0.047 3	0.032 7	0	0	0.040 4
t_2	0.531 7	0.534 5	0.531 7	0.021 78	0	0	0.377 1
t_3	0.362 4	0.339 7	0.345 8	0.013 21	0	0	0.249 2
t_4	0.574	0.520 6	0.537 2	0.008 7	0	0.4	0.387 3

分析得到的结果,可以看出同一主机,当检测发现攻击的传感器越多时,说明攻击发生的概率越大,相应的主机安全状况也越严重。

不考虑时间参数,对网络安全态势值进行计算,得到的实验计算结果如表 4 所示。

表 4 不考虑时间参数的实验计算结果

time	mill	locke	pascal	hume	robin	www. af. mil	SA
WH	0.34	0.18	0.18	0.2	0.05	0.05	1
t_0	0.036 7	0.04	0.036 7	0.04	0	0	0.034 3
t_1	0.045 9	0.05	0.045 9	0	0	0	0.032 9
t_2	0.5	0.5	0.5	0	0	0	0.35
t_3	0.2	0.175	0.183 4	0	0	0	0.132 5
t_4	0.4	0.35	0.366 8	0	0	0.4	0.285

绘制两种计算方法的网络安全态势值曲线,如图 3 所示。从曲线图可以看出,在 t_3、t_4 时刻两次实验结果有着明显的不同,根据实际情况分析,同一次攻击场景中,从 t_2 到 t_3,攻击仍在继续,虽然暂时检测不到某些攻击,但从 t_4 的态势值可以看出,攻击只是进行了蛰伏,网络安全态势值不应出现陡降的情况,两种模型都出现了态势值下降的情况,但本文模型明显比不带时间参数的模型的态势值下降幅度低。分析主机安全态

势值计算结果,发现不考虑时间参数的模型中,直接忽略了 t_3 时刻前面的攻击阶段,mill、locke、pascal 三个主机的态势值的计算结果均比较小。整个攻击阶段,攻击虽然时断时续,但是整体的态势应该呈现连续的状态,由于不考虑时间的影响因素,使得评估结果出现了明显误差,评估没有连续性,同时也缺乏实时性。

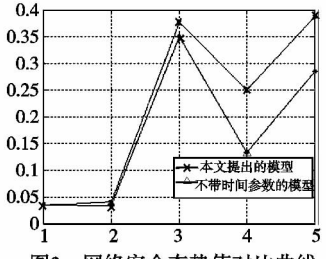

图3　网络安全态势值对比曲线

本文提出的基于时变 D-S 证据理论的网络安全态势评估模型,对 D-S 证据理论进行了扩展,引入了时变性的概念,提出用带时间参数的函数来描述 m 值的变化规律,使得改进后的 m 值由静态值变为随时间衰减函数变化的动态值,证据信息更具备合理性和通用性。在此基础上,根据层次化模型对网络安全态势进行量化,得到网络安全态势值。通过曲线图对比,本文模型更符合实验网络情况,得到的结果更为准确。因此,采用时变 D-S 证据理论对网络安全态势进行评估的层次化模型能够提高网络安全态势评估的准确性和连续性。

4　结束语

本文针对 D-S 证据理论中证据会随着时间的推移而变化的问题,对 D-S 证据理论进行了扩展,提出了时变的 D-S 证据理论,对证据进行了完善,使得证据信息更符合实际情况。使用改进的时变 D-S 证据理论方法对多传感器的证据进行融合,得到威胁的发生概率;其次,提出带有时间参数的层次化模型,融合计算网络安全态势值,实现了网络安全态势的量化评估。最后,将本文提出的模型和不带时间参数的模型进行了对比实验,可以看出采用基于时变 D-S 证据理论的网络安全态势评估的模型相对于不考虑时间参数的评估模型,有效地提高了网络安全态势的评估的准确性和连续性。

参考文献:

[1] BASS T. Intrusion detection systems & multisensor data fusion :creating cyberspace situational awareness[J]. Communications of the ACM. 2000,43(4):99-105.

[2] BASS T. Multi-sensor data fusion for next generation distributed intrusion detection systems[C]//Proc of IRIS National Symposium on Sensor and Data Fusion. 1999:24-27.

[3] SHIFFLETS J. A technique independent fusion model for network intrusion detection[EB/OL]. (2005-10-13). http://www. jcu. edu/math/swarm/papers/Jason 2005. pdf.

[4] D'AMBROSIO B. Security situation assessment and response evaluation (SSARE)[C]//Proc of DARPA Information Survivability Conference & Exposition II. Washington DC: IEEE Computer Society, 2001:387-394.

[5] LAKKARAJU K, YURCIK W, LEE A J. NVision IP: netflow visualizations of system state for security situational awareness[C]//Proc of ACM Workshop on Visualization and Data Mining for Computer Security. New York: ACM Press, 2004:65-72.

[6] YIN Xiao-xin, YURCIK W, SLAGELL A. The design of VisFlowConnect-IP: a link analysis system for IP security situational awareness [C]//Proc of the 3rd IEEE International Workshop on Information Assurance. Maryland:IEEE Press,2005:141-153.

[7] SHAFER G. A mathematical theory of evidence[M]. Princeton: Princeton University Press, 1976.

[8] 陈秀真,郑庆华,管晓宏,等. 层次化网络安全威胁态势量化评估方法[J]. 软件学报,2006 ,17(4): 885-897.

[9] 孙全,叶秀清,顾伟康. 一种新的基于证据理论的合成公式[J]. 电子学报,2000,28(8): 117-119.

[10] 韦勇,连一峰,冯登国. 基于信息融合的网络安全态势评估模型[J]. 计算机研究与发展,2009,46(3):353-362.

[11] The snort project. Snort^R users manual 2. 8. 5[RB/OL]. 2009-10-22. http://www. snort. org/assets/125/snort_manual-2_8_5_1. pdf.

[12] 史超,程咏梅. 基于证据冲突度的多传感器冲突信息组合方法[J]. 计算机应用研究,2011,28(3):865-868.

[13] 张勇,谭小彬,崔孝林,等. 基于 Markov 博弈模型的网络安全态势感知方法[J]. 软件学报,2011,22(3):495-508.