层次化网络信息内容安全事件态势评估模型

葛 琳,季新生,江 涛

(国家数字交换系统工程技术研究中心,郑州 450002)

摘 要:针对网络中信息内容安全事件的态势评估问题,通过对网络信息内容安全事件的多维特征分析,提出了一种层次化的信息内容安全事件态势评估模型及参数计算方法。该模型采用层次式结构,分别对事件级、区域级和系统级的态势评估值进行计算。其中,事件级态势利用事件特征中的行为特征和内容特征进行计算;区域级态势则依据关系特征和位置特征;系统级态势整合所涉及的各区域级态势。对各级态势评估值参数的计算方法进行了定义。仿真实验结果表明:该模型及计算方法具有可行性和可靠性,在对信息内容安全事件的态势评估过程中能够有效反映事件的影响程度并把握其变化规律。

关键词:信息处理技术;信息内容安全事件;态势评估;层次化

中图分类号:TP393 **文献标志码**:A **文章编号**:1671-5497(2016)02-0556-12

DOI:10.13229/j.cnki.jdxbgxb201602034

Hierarchical situation evaluation model for network information content security incidents

GE Lin, JI Xin-Sheng, JIANG Tao

(*National Digital Switching System Engineering and Technological Research Center, Zhengzhou* 450002, *China*)

Abstract: To solve the problem of situation evaluation for network information content security incidents, a hierarchical situation evaluation model parameter calculation method are proposed, through the analysis on the multi-dimension characteristics of network information content security incidents. The model uses hierarchical structure and calculates the situation assessment values of incident level, area level and system level respectively. Among these levels, the incident level is based on the behavioral characteristics and content characteristics of the incident characteristics; the area level is according to the relationship characteristics and location characteristics, and system level integrates the involved area levels. The method to calculate the parameters is defined. Simulation results show that the model and method is feasible and reliable. It can reflect the impaction of incidents effectively and grasp the change rules of the incidents in situation evaluation of information content security incidents.

Key words: information processing technology; information content security incidents; situation evaluation; hierarchical

收稿日期:2014-02-27.

基金项目:"863"国家高技术研究发展计划项目(2011AA010605);国家自然科学基金创新群体项目(61521003);国家科技重大专项项目(2012ZX03006002-010,2013ZX03006002).

作者简介:葛琳(1978-),女,博士研究生.研究方向:网络信息安全.E-mail:lingesnow@126.com

通信作者:江涛(1974-),男,副研究员.研究方向:移动互联网安全.E-mail:jiangtao@outlook.com

0 引　言

信息内容安全事件（Information content security incidents，ICSI）指利用信息网络发布、传播危害国家安全、社会稳定和公共利益内容的安全事件[1-2]。目前，信息内容安全领域的研究主要有：针对以传输特定信息为目的的信息渗透的检测技术研究[3]、针对网络信息内容安全的控制模型及其评估的研究[4]、基于文本内容的事件分类技术[5]以及通过对多媒体内容的识别，发现其中隐藏的安全事件[6]等。

现有的态势评估技术主要针对攻击类的网络安全事件，对具有合法网络用户身份却发送不良信息内容所构成的安全事件进行态势评估的研究极少。近年来的态势评估技术主要有：Roesch[7]根据攻击的破坏程度和目标漏洞的风险级别来对报警数据划分威胁等级，实现对安全态势的评估；Porras 等[8]提出了基于受系统任务影响的预警优先级评估方法；Hariri 等[9]针对通信协议和网络设备的脆弱性，提出了数据传输速率、缓冲区占用率等网络性能指标，通过大量分布在网络中的 agent 获取与脆弱性相关的信息，对网络的安全状况进行评估；文献[10]提出了一种基于 IDS 海量预警信息和网络性能指标的，并结合服务、主机本身的重要性及网络系统的组织结构的层次化网络安全威胁态势量化的评估方法；文献[11]使用隐马尔可夫模型描述主机的各种安全状态和转换关系，通过安全状态之间的转换概率评估主机受到的威胁，而网络所处的威胁等级则由各主机的威胁等级综合决定；文献[12]提出基于信息融合的网络安全态势评估模型，包含数据源融合、态势要素融合和节点态势融合 3 种算法；文献[13]针对网络安全中涉及的动态性因素，如安全漏洞和安全威胁等，提出了一个安全指标框架，对各安全风险因素进行量化。

相较于攻击类事件，ICSI 以传播不良信息内容为目的，强度小、信息繁杂、具有隐蔽性，不会像攻击类事件一样破坏网络中硬件设备导致故障和瘫痪等，也不存在类似 IDS 日志的告警信息。因此，对该类事件进行态势评估的难度较大。如何通过对攻击类网络安全事件的态势评估技术的学习和借鉴，实现对信息安全事件的态势评估，是本文要解决的主要问题。通过对已有相关研究的分析可以看出，对网络中 ICSI 进行评估需要考虑以下几点：①评估网络或系统的组织架构；②涉及节点的重要性程度或等级的先验知识；③事件的特征及其对态势评估值的影响。通过上述分析，本文对 ICSI 的特征进行了分析，提取其所具有的多维特征，提出了一种层次化评估模型和参数计算方法，对 ICSI 进行态势评估。该模型采用层次式结构，将复杂的问题进行分步解决。首先，将各组成维度特征按照一定的关系进行分组；然后，按照关系构成层次化的评估结构体系；最后，计算各层的评估结果，并整合汇总。实验结果表明，本文提出的模型和方法具有可行性和可靠性，在对 ICSI 的态势评估中切实有效。

1 网络信息内容安全事件特征分析

数据集的主要来源为 VAST 2008 中的 Cell Phone Social Network 数据集[14]和 Enron 公司 2009 年邮件数据集[15]。其中，Cell Phone Social Network 数据集包含了 400 人 10 天共计 9834 条通信记录。如表 1 所示，From 和 To 分别表示通信双方的号码编号，Date time 表示按照需求划分出的时间段编号，Duration 为通信时长，Type 为通信类型，Cell Tower 为主叫用户所属的基站编号。Enron 公司邮件数据集中，采用 87 448 个用户的 255 636 封邮件数据。如表 2 所示，Domain Name-i 和 Domain Name-j 分别表示邮件双方的地址域名，Time 为邮件发送时间，Subject 为邮件的主题。

表 1　VAST 2008 Cell Phone Social Network 数据集内容及格式

Table 1　Content and format of VAST 2008 Cell Phone Social Network

From	Date time	Duration/s	Type	To	Cell Tower
349	20060601 0008	1634	Calling	23	1
379	20060601 0011	640	Calling	364	24
392	20060601 0012	1182	Calling	27	29
17	20060601 0014	578	Calling	339	23
272	20060601 0015	887	Calling	251	29
⋮	⋮	⋮		⋮	⋮
118	20060610 2356	2356	Calling	213	6

表2 Enron邮件数据集内容及格式

Table 2 Content and format of mail data sets in Enron

From	Domain Name-i	To	Domain Name-j	Time	Subject
1497	enron. com	1498	mailman. enron. com	2001-07-11 08:29:22	Confidential Information and Securities Trading
1497	enron. com	409	enron. com	2001-07-11 08:29:22	Confidential Information and Securities Trading
60	enron. com	115	enron. com	2001-05-11 15:26:17	Null
1490	enron. com	4652	enron. com	2001-04-17 14:39:00	Expertfinder—The Power of Who
49879	mcnallytemple. com	1162	eslawfirm. com	2001-05-22 03:01:00	IEP News 5/22
801	enron. com	3164	bracepatt. com	2001-06-03 18:26:00	Amendment of the CPUC subpoena for production on June 4 with mont hly updates, and clarification on confidential treatment
1238	caiso. com	3304	caiso. com	2000-07-12 02:26:00	CAISO Notice — Congestion Reform Proposal — Apendix B
36	enron. com	4001	brobeck. com	2001-06-06 16:48:00	ISO's Response to BPA Rebuttal of Sheffrin Study—Confidential Atty Client work product
49879	mcnallytemple. com	4005	iepa. com	2001-04-19 03:45:00	IEP in the News, and other headlines
⋮	⋮	⋮	⋮	⋮	⋮
1490	enron. com	5827	onlinemailbox. net	2001-07-17 06:42:00	Re: Public Policy Contacts for California

网络通信数据集可以反映网络中用户的多个通信特征,通过对通信特征的分析可以找到其中蕴含的规律,对其进行归纳可折射出网络中各类安全事件的状况。通过对上述两个数据集的分析可以得到:

(1)行为特征(Behavioral character)

表1中,大部分用户的通信时长在一个常规的阈值范围之内,而极少数用户的通信时长超过或低于该阈值,累积超长或超短通信次数较多的地址或号码具有一定的非常规性。表2中,大部分用户的邮件内容不具有重复性,即具有不同的主题,那些同一通信类型且具有相同内容类似广播的通信通常具有非常规性。同时,大部分用户某时段内的通信次数通常保持在一个常规的阈值范围之内,而极少数用户的通信次数超过了该阈值,表明该号码或地址具有特殊用途,例如,作为商业联络或者为ICSI的发送源,前者属于合法通信行为,后者则需结合通信内容做出判别。

(2)关系特征(Relation character)

表1和表2中,大多数用户的通信对象较为固定,即仅与一定范围内的人进行联络,符合人类社交的群体性特征,而那些通信对象数目过多的用户可能具有特殊的用途。

(3)位置特征(Location character)

表1中,大部分用户在一段时间内通信所涉及的基站数目较少,符合用户活动范围的有限区域性特征。极少数用户涉及的基站范围较多,表明该用户具有一定的特殊性。同理,表2中,大部分用户在一定时间段内的IP地址是较为固定的,极少数用户的位置变化频繁,表明了该用户具有特殊性。

(4)内容特征(Content character)

对表1和表2中用户的行为特征、关系特征和位置特征进行关联分析,可以得到网络中特殊用户的地址或号码,但如果需要进一步明确ICSI事件的具体内容,则需有针对性地对此类用户的具体通信内容进行分析。

通过上述分析可知,ICSI具有多维特征,如图1所示。同时,时间(When)、地点(Where)、人物(Who)和内容(What)是能够清楚描述一个事件的四要素,在对ICSI的态势评估中,需要充分考虑到这4个因素。其中,When是事件的发生

区间即通信时间，Where 可在用户的位置特征中得到反映，Who 可在表示用户的通信关系特征信息中获取，What 则由用户的行为特征和通信内容特征决定。从图中可以看出，在网络多维通信数据集中，隐含了各类安全事件发生的要素，如何充分利用此类信息，整合 ICSI 所涉及的各维数据，是实现对 ICSI 全面的态势评估的基础。下面将针对此问题，提出一种层次化的评估模型，并对其各指数的量化方法进行论述。

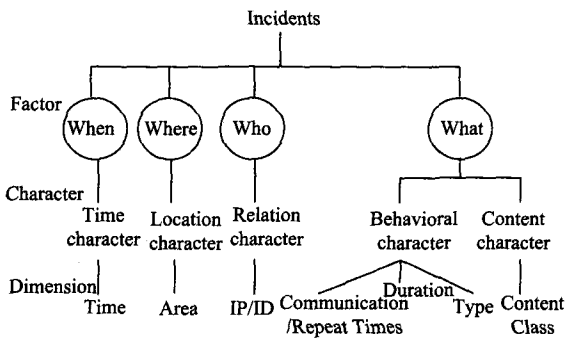

图 1　ICSI 的多维特征组成

Fig. 1　Multi-dimension characteristics of ICSI

2　信息内容安全事件态势评估体系

通过前文的分析可知，网络 ICSI 的态势通过多维通信信息展示。首先，对通信数据分析得到 ICSI 的事件的自身信息。例如，通过对通联关系的分析可确定事件的影响（发送次数、涉及人数）；通过对通信类型的分析可确定事件的类型（语音、视频、邮件等）；通过对主题、关键词等内容的分析可以确定内容类别（政治类、经济类、军事类等）；通过对涉及的通信地址的分析可得出事件的目标（目标人群、地区）。其次，通过对各个区域内涉及 ICSI 的地址或号码的分析可得出不同地区的态势指数；最后，综合各个区域的 ICSI 态势指数得出全网的 ICSI 态势值。

2.1　层次化评估模型

根据对事件的分析过程，本文提出了一个层次化 ICSI 态势评估模型，如图 2 所示。该模型分为数据层、事件层、区域层和系统层 4 个层次，采用自下而上，先局部后整体，先判别事件后根据事件关联的方法对网络中 ICSI 的态势进行评估。图 2 中，在数据层输入相关的 ICSI 通信记录；在事件层利用行为特征中的通信时长、通信次数、通信类型和内容特征中的内容类别进行事件级态势指数的评估；根据位置特征和关系特征中涉及到的用户位置和地址/号码等相关信息确定事件所属的区域，结合事件层得出的事件级态势指数，计算区域层态势指数；系统层整合各个区域的态势指数给出整个网络的态势值。

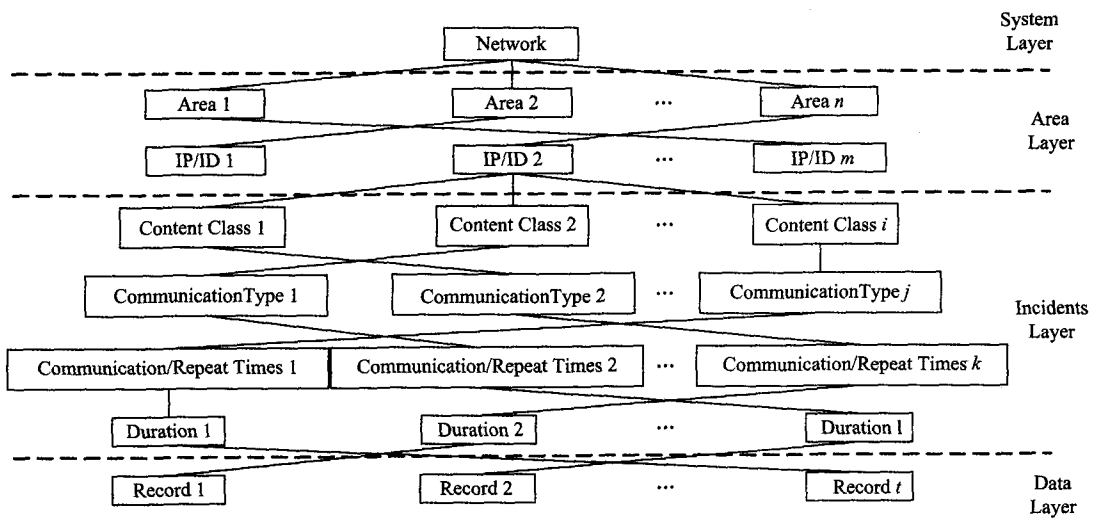

图 2　层次化 ICSI 态势评估模型

Fig. 2　Hierarchical situation evaluation model of ICSI

定义 1　通信记录 R(Record)。引发 ICSI 的网络通信记录，表示为：$R = \{Time, Duration, Times, Type, Class, SIP/ID, DIP/ID, Area\}$。其中，$Time$ 为研究的时间区间，$Duration$ 为通信时长，$Times$ 为通信次数，$Type$ 为通信类型，$Class$ 为内容类别，SIP/ID 为源地址 / 号码，DIP/ID 为目标地址 / 号码，$Area$ 为用户所属的基站/域名位置。

定义2 事件级态势指数 IF(Incident factor)。表示ICSI发生时对整个网络ICSI的影响程度。通过ICSI中的通信类型和内容类别结合通信时长、通信次数对ICSI的影响,在事件层做出判断。

定义3 区域级态势指数 AF(Area factor)。表示ICSI事件所涉及区域的态势指数。综合本区域中用户及其涉及事件的态势指数给出 AF。

定义4 系统级态势指数 SF(System factor)。表示整个通信系统中ICSI的态势总指数。整合各个区域的 AF,给出整个系统的 SF。

2.2 评估指数的量化计算

2.2.1 事件级

事件级态势指数的计算包括了对事件通信次数、通信时长、通信类型和内容类别的综合衡量,IF_i 的值越大,表示事件级态势指数对整个系统的ICSI态势指数的影响越大。为了更加真实地反映网络中ICSI的变化,本文将一天划分为 h 个时间段。对于给定的分析时间窗口 Δt,定义 t 时刻事件 i 的态势指数为:

$$IF_i(t) = f(\vec{\theta},\vec{l},\vec{N}_i(t),\vec{M}_i(t)\vec{D}_i(t)) = [\vec{\theta}\cdot\vec{N}_i(t)+\vec{l}\cdot\vec{M}_i(t)]\cdot 10^{\vec{D}_i(t)} \quad (1)$$

(1) $\vec{\theta}=(\theta_1,\theta_2,\cdots,\theta_h)$ 为通信次数向量。$\vec{\theta}$ 的初始值可根据不同时段的网络中的正常通信次数 $T_k(k=1,\cdots,h)$ 进行赋值。采用1~5五个等级对其进行描述:非常少,少,正常,多,很多。取值越大则表示通信次数越多。然后,对此进行归一化处理,得到 $\vec{\theta}$ 的元素值,即:

$$\theta_k = \frac{T_k}{\sum_{t=1}^{h} T_t} \quad (2)$$

(2) $\vec{l}=(l_1,l_2,\cdots,l_h)$ 为通信时长向量。$\vec{l}$ 的初始值可根据不同时段网络中平均通信时长 $D_l(l=1,\cdots,h)$ 进行赋值。与 $\vec{\theta}$ 的赋值类似,也可采用1~5五个等级对其进行描述:非常短,短,正常,长,很长。取值越大则表示通信时长越长。然后,对此进行归一化处理,可得到 $\vec{l}$ 的元素值:

$$l_l = \frac{D_l}{\sum_{t=1}^{h} D_t} \quad (3)$$

(3) $\vec{N}_i(t)$ 和 $\vec{M}_i(t)$ 分别为 t 时刻ICSI事件的行为特征中的超长通话/超短通话和频繁通信次数向量。其中,$\vec{N}_{ij}(t)=(N_{ij1},N_{ij2},\cdots,N_{iju})$ 为第 j 个时间段内从 t 至 $t+\Delta t$ 频繁通信此类维度特征 i 的发生次数与该时间段内的通信总次数的比值;$\vec{M}_{ij}(t)=(M_{ij1},M_{ij2},\cdots,M_{iju})$ 为第 j 个时间段内从 t 至 $t+\Delta t$ 超长通话/超短通话此类维度特征 i 的发生次数与该时间段内的通信总次数的比值,u 为 Δt 时间段内维度特征 i 的通信类型数目。

(4) $\vec{D}_i(t)$ 为事件严重程度向量。$\vec{D}_{ij}(t)=(D_{ij1},D_{ij2},\cdots,D_{iju})$ 为第 j 个时间段内从 t 至 $t+\Delta t$ 维度特征 i 对ICSI态势影响的程度。其中,$j=1,\cdots,h$。u 为 Δt 时间段内维度特征 i 的通信类型数目。u 和 $\vec{N}_{ij}$ 的取值可通过对数据层的通信记录数据集统计得到。$\vec{D}_{ij}$ 作为一种程度判别指数,其值的确定受多种因素影响,逻辑关系复杂,本文结合客观统计信息和主观经验知识,即按照ICSI的定义对内容进行分类得到内容类别,同时结合通信类型对用户的影响。例如,将通信内容对用户的影响程度排序为:特殊关注事件>政治类>军事类>社会类>经济类等,按照通信类型对用户的影响程度从大到小排序为:视频类>语音类>短信类>邮件类>其他类,$\vec{D}_{ij}$ 值的判断原则如表3所示。

表3 $\vec{D}_{ij}$ 的判断原则

Table 3 Judgment principle of $\vec{D}_{ij}$

类型	Special	Political	Military	Social	Economic
Video	10	9	8	7	6
Audio	9	8	7	6	5
Message	8	7	6	5	4
E-mail	7	6	5	4	3
Else	6	5	4	3	2

(5) 定义运算 $10^{\vec{D}_{ij}}=(10^{D_{ij1}},10^{D_{ij2}},\cdots,10^{D_{ijn}})$。本文对不同等级的ICSI的等效性进行了试验,大多数情况下,100次程度为1的事件态势指数与10次程度为2的事件态势指数、1次程度为3的事件态势指数基本等效。例如,根据实际情况和经验,3次程度为2的事件态势指数比2次程度为3的事件态势指数要小,若按照 $\vec{N}_{ij}\cdot\vec{D}_{ij}$ 直接进行计算则为 $3\times2=6=2\times3$,二者相等,与实际情况和直观判断不相符。若按照 $\vec{N}_{ij}\cdot$

$10^{\vec{p}_{ij}}$ 计算则为 $3\times10^2<2\times10^3$，与实际情况和直观判断接近。

2.2.2　区域级和系统级

本文对区域级和系统级态势指数的设计采用相同的原理。t 时刻区域 n 的区域级态势指数 AF_n 越大，说明该区域的 ICSI 事件态势状态越严重。

$$AF_n(t)=f(\vec{p},\vec{IF}(t))=\vec{p}\cdot\vec{IF}(t) \quad (4)$$

(1) $\vec{p}=(p_1,p_2,\cdots,p_m)$ 为事件在区域内涉及的所有用户所占的重要程度权重向量，m 为区域中的用户数目。可以根据网络安全管理的实际需求进行预先设定 $I_w(w=1,2,\cdots,m)$，分别用 1～3 等级对应低、中、高来表示其重要性，并采用归一化处理：

$$p_w=I_w\Big/\sum_{t=1}^{m}I_t \quad (5)$$

(2) $\vec{IF}(t)=(IF_1,IF_2,\cdots,IF_m)$ 为 t 时刻区域 n 中用户涉及的事件态势向量，其中 IF_1、IF_2、…、IF_m 为根据式(1)计算得出的事件级态势指数。

与区域级态势指数的计算方法类似，t 时刻系统级 ICSI 态势指数 $SF(t)$ 越大，表明该系统中的 ICSI 发生状况越严重，越值得引起关注和重视。

$$SF(t)=f(\vec{q},\vec{AF}(t))=\vec{q}\cdot\vec{AF}(t) \quad (6)$$

(3) $\vec{q}=(q_1,q_2,\cdots,q_n)$ 为组成系统的各区域重要性权重向量，n 为系统所包含的区域数目。预先设定各个区域的重要性权重，并采用归一化处理 $(v=1,2,\cdots,n)$。

$$q_v=Z_v\Big/\sum_{t=1}^{n}Z_t \quad (7)$$

(4) $\vec{AF}(t)=(AF_1,AF_2,\cdots,AF_n)$ 为 t 时刻系统中的各区域态势向量，其中 AF_1、AF_2、…、AF_n 为根据式(4)计算得出的区域级态势指数。

3　实验分析

为全面验证层次化 ICSI 态势评估模型的有效性和可靠性，本文实验分为两部分进行。实验一，利用开源数据集，采用较大的时间窗口，进行事件低维度特征的粗粒度态势评估；实验二，建立局域网仿真环境，采用较小的事件窗口，结合事件多维度特征进行细粒度的态势评估。

3.1　实验一

本节实验采用 VAST 2008 中的 Cell Phone Social Network 数据集和 Enron 公司 2009 年邮件数据集作为实验数据。其中，采用 VAST 2008 的数据集中的 2006 年 6 月 1 日至 5 日的数据进行态势评估，将超长通话(Long duration)、超短通话(Short duration)和频繁通信(Frequent communication)作为事件的行为特征，如图 3 所示。根据数据集的具体数据和实验需求，图 3 中选取基站编号为 10 的基站下用户作为研究对象，在选取的 5 天时间段内，最长通信时长为 1732 s，最短通信时长为 166 s，平均通信时长为 1030 s，平均通信次数为 5。根据先验知识和实验数据的选取便利，将大于 n 的通信记为超短通话，认为大于 10 次的通信为频繁通信(Frequent communication)。采用 Enron 公司 2009 年邮件数据集中的 2001 年 5 月 29 日至 6 月 27 日的数据进行态势评估，将频繁通信作为事件的行为特征。如图 4 所示，选取域名为 mcnallytemple. com 和 enron. com 下的用户作为研究对象，在选取的 30 天时间段内，平均通信次数为 56，将大于 112 次的通信记为通信次数较多。

设定 Δt 为 1 天，划分出 4 个时间段 T_k 分别

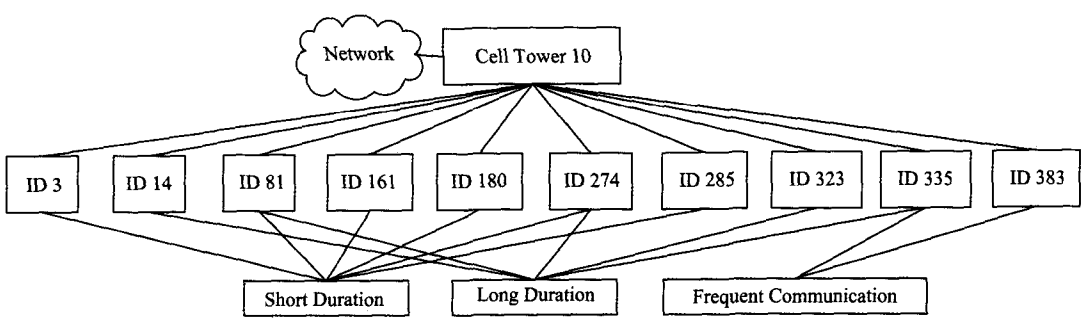

图 3　VAST 2008 数据集下的 ICSI 态势评估模型

Fig. 3　Situation evaluation model of ICSI in VAST 2008

赋值为1、2、3、4,来表示通信次数的变化情况:非常低,中,中,高,对其进行归一化处理后得 $\vec{\theta}$ = (0.091,0.273,0.273,0.364)。事件特征和涉及的ID/IP及其重要性如表4和表5所示,对Enron 2009中两域名的权重赋值分别为0.6和0.4。

利用前文介绍的层次式ICSI态势评估模型

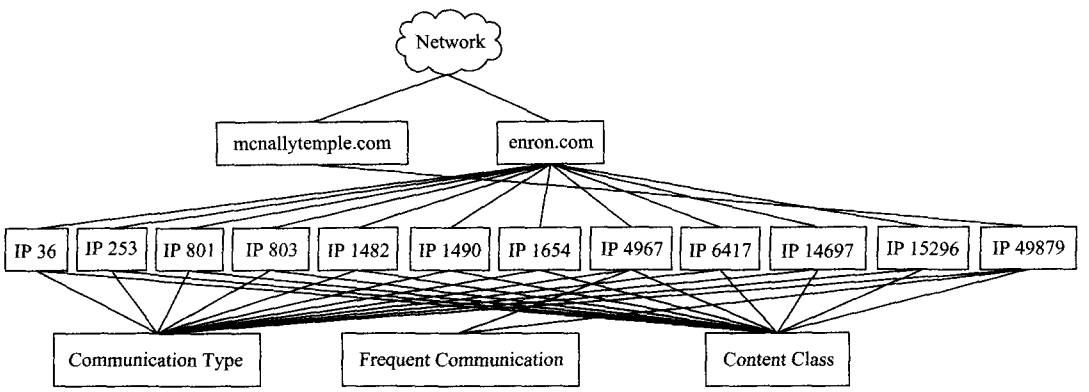

图4 Enron 2009邮件数据集下的ICSI态势评估模型

Fig. 4 Situation evaluation model of ICSI in Enron 2009

表4 VAST 2008数据集下事件特征、ID及其重要性

Table 4 Incidents characteristics、ID and importance in VAST 2008

ID Number	Incidents Characteristics	Degree of Incidents Characteristics	Weights of Incidents Characteristics	Degree of ID importance	Weights of ID importance
3	{Short Duration}	{3}	(1)	38	0.104
14	{Long Duration}	{3}	(1)	43	0.117
81	{ Short Duration; Long Duration}	{3; 3}	(0.5; 0.5)	53	0.145
161	{ Short Duration}	{3}	(1)	19	0.052
180	{ Short Duration}	{3}	(1)	38	0.104
274	{ Short Duration; Long Duration}	{3; 3}	(0.5; 0.5)	22	0.060
285	{ Short Duration}	{3}	(1)	43	0.117
323	{ Long Duration}	{3}	(1)	38	0.104
335	{Long Duration; Frequent Communication}	{3; 2}	(0.6; 0.4)	53	0.145
383	{Frequent Communication}	{2}	(1)	19	0.052

表5 Enron 2009邮件数据集下事件特征、IP及其重要性

Table 5 Incidents characteristics、IP and importance in Enron 2009

IP Number	Incidents Characteristics	Degree of Incidents Characteristics	Weights of Incidents Characteristics	Degree of ID importance	Weights of ID importance
36	{Communication Type; Content Class}	{3; 3}	(0.5; 0.5)	29	0.070
253	{Communication Type; Content Class}	{3; 3}	(0.5; 0.5)	36	0.087
801	{Communication Type; Content Class}	{3; 2}	(0.6; 0.4)	42	0.101
803	{Communication Type; Content Class}	{3; 2}	(0.6; 0.4)	42	0.101
1482	{Communication Type; Content Class}	{3; 2}	(0.6; 0.4)	42	0.101
1490	{Communication Type; Content Class}	{3; 2}	(0.6; 0.4)	36	0.087
1654	{Communication Type; Content Class}	{3; 1}	(0.75; 0.25)	18	0.043
4967	{Communication Type; Frequent Communication; Content Class}	{3; 2; 1}	(0.5; 0.33; 0.17)	53	0.127
6417	{Communication Type; Content Class}	{3; 1}	(0.75; 0.25)	18	0.043
14697	{Communication Type; Content Class}	{3; 2}	(0.6; 0.4)	18	0.043
15296	{Communication Type; Content Class}	{3; 1}	(0.75; 0.25)	29	0.070
49879	{Communication Type; Frequent Communication; Content Class}	{3; 2; 3}	(0.375; 0.25; 0.375)	53	0.127

的计算方法，对图 3 和图 4 中的模型进行分析，结合表 4 和表 5 中的数据，对纵坐标做归一化处理，可得到如下实验结果：

(1) VAST 2008 和 Enron 2009 数据集下的 ICSI 事件级态势(分别以 ID335 和 IP4967 为例)。

图 5 中，6 月 1 日和 6 月 4 日分别出现了两个特征的峰值，图 6 中，在 6 月 6 日出现态势的最高峰点，6 月 20 日出现态势的较小峰值点，其他时间段基本处于零值点。以 Enron 2009 数据集中用户 IP4967 为例进行说明：在提取的时间段内，通过与数据集中的数据对照发现，其邮件发送数目在 6 月 6 日为 413，6 月 20 日为 120，均超过了设定的阈值，且 6 日的次数远大于 20 日的次数，而在 5 月 29 日至 6 月 27 日的其他时间均没有出现超过阈值的通信，这一行为特征从图中得到了良好的反映。同时，从图 6 中还可以看出，通信类型和内容类别特征权重的赋值对于用户 IP4967 来说是不变的，但是随着频繁通信特征的态势变化，二者也随之发生了变化。实际中，当某用户采用固定的通信类型进行频繁通信时，即使其传递的消息内容权重值始终不高，如固定内容的垃圾邮件，仍应引起网络管理员的重视。由此，

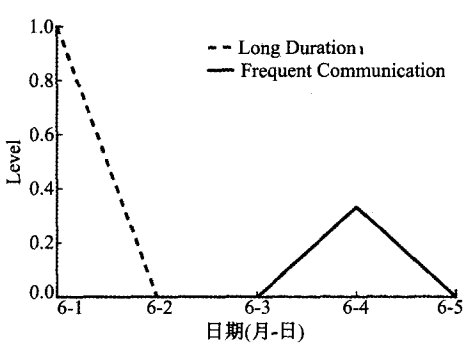

图 5　VAST 2008 数据集中 ID335 的 ICSI 事件级态势

Fig. 5　ICSI incidents situation of ID 335 in VAST 2008

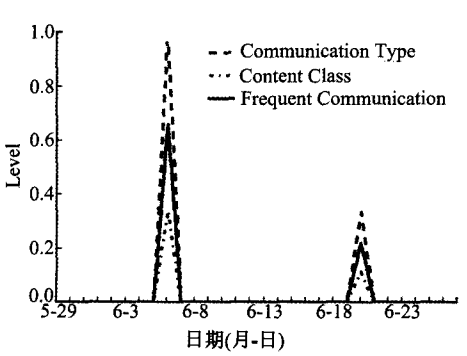

图 6　Enron 2009 数据集中 IP4967 的 ICSI 事件级态势

Fig. 6　ICSI incidents situation of IP 4967 in Enron 2009

ICSI 的各个维度特征之间可以产生相互影响，且本文中对事件级态势评估定义的计算方法是有效的。

(2) VAST 2008 和 Enron 2009 数据集下的 ICSI 区域级态势(分别以 ID161、285、323 和 335；IP253、801、1654 和 4967 为例)。

图 7 和图 8 为区域内各 ID 和 IP 的态势评估变化。以图 7 中的 VAST 2008 为例进行说明：ID335 的事件级态势如图 5 所示，在区域级态势计算时加入了该用户的重要性权重 0.145，如图 7 所示，其态势评估值的大小较事件级态势下降了一些，原因为与区域内的其他用户分配了权重(即对整体态势的影响程度)，但没有对该用户个体的态势变化趋势产生影响。因此，加入用户重要性权重之后，可以更加突出定义的重要用户的 ICSI 态势变化。对于其中重要性赋值较低的用户，其变化仍为研究的对象，但不会对整体态势产生大的影响，符合网络中的实际运行状况和信息安全管理的需求。

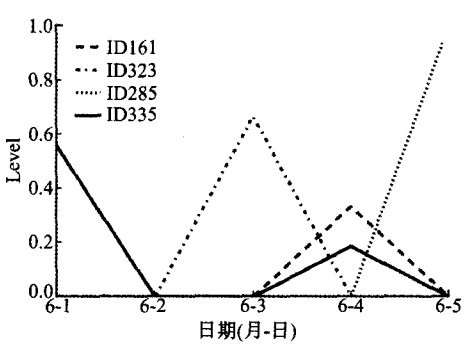

图 7　VAST 2008 数据集中 ID161、285、323 和 335 的 ICSI 区域级态势

Fig. 7　ICSI area situation of ID161,285,323 and 335 in VAST 2008

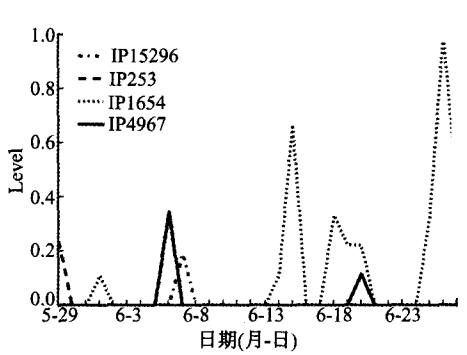

图 8　Enron 2009 数据集中 IP253、1654、4967 和 15296 的 ICSI 区域级态势

Fig. 8　ICSI area situation of IP 253,1654,4967 and 15296 in Enron 2009

(3)VAST 2008 和 Enron 2009 数据集下的 ICSI 系统级态势

整合系统中所涉及的各区域级态势评估值即可得到系统整体的态势评估值变化。本节实验中,VAST 2008 只涉及一个基站,Enron 2009 涉及两个域名。将图 9 和图 10 与前文中的事件级、区域级态势图进行对比可以看出,二者的系统级态势图中均包含了个体和区域的态势走势特点。其中,VAST 2008 基站 10 下的 ICSI 态势变化趋势较为连续,平缓处较多;Enron 2009 两域名下的 ICSI 态势变化趋势起伏较大。结合数据库中的数据分析可知,电信网络中 ICSI 各特征的变化较为显著,在不同时段虽然表现的特征维度不同,但均具有一定的特征变化。在互联网络中,ICSI 各特征的变化则具有一定的突发性。由此可以看出,本文提出的模型和参数计算方法,结合了网络构成的各元素,从最基本的各用户行为特征等处着手,对 ICSI 的态势变化具有敏感性,能够有效把握其变化趋势。

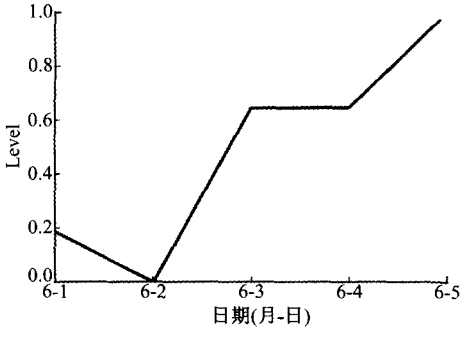

图 9 VAST 2008 数据集下的 ICSI 系统级态势

Fig. 9 ICSI system situation in VAST 2008

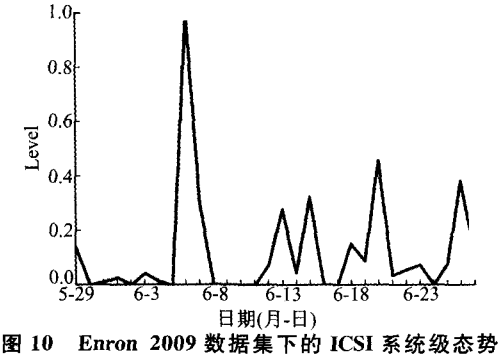

图 10 Enron 2009 数据集下的 ICSI 系统级态势

Fig. 10 ICSI system situation in Enron 2009

3.2 实验二

实验一中的数据来源为开源数据集,可供选择的特征维度较低,选择的时间窗口较大。本节将通过自建局域网仿真环境模拟产生 ICSI 事件,对涵盖 5 个特征维度的 ICSI 进行细粒度的态势评估。该局域网网段为 192.168.1.0～24,其中 192.168.0.21、192.168.0.22 和 192.168.0.23 分别为 3 个区域地址,系统地址为 192.168.0.24,三个区域的重要性权重赋值分别为 0.3、0.4 和 0.3。实验中通过计算机终端通信软件模拟产生超长通话、超短通话和频繁通信,涉及的通信类型为视频、音频和信息,内容类别设定为军事、政治、经济和特殊类。基于层次化的 ICSI 态势评估模型如图 11 所示,各事件特征重要性和用户个体权重信息的设置方法与实验一中相同,由于涉及终端用户数目较多,在此不再赘述。

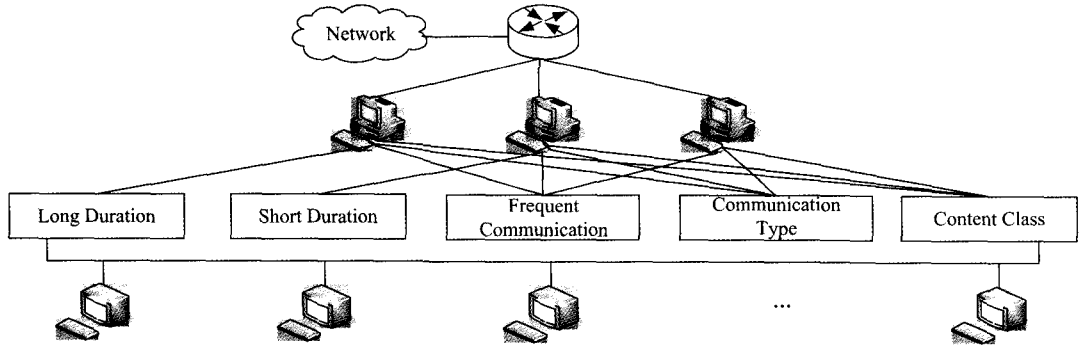

图 11 局域网仿真环境下的 ICSI 态势评估模型

Fig. 11 ICSI situation evaluation model in LAN

实验中,产生的通信数据总数目为 166 条/min,其中,出现超长通话、超短通话和频繁通信的条数/分钟,出现的比例分别为 0.5%、0.5%和 0.5%。将超过 600 s 的通信时长记为超长通信,低于 10 s 的通信时长记为超短通信,超过 6 次/min 的通信次数记为频繁通信。各通信类型占总通信类型的比例设置为:视频 10%、音频 50%和信息 40%,按照通信类型对用户的影响

程度从大到小设置为，视频类＞语音类＞短信类＞邮件类等。同时，为了凸显态势评估模型对特殊类事件的有效感知能力，将各内容类别占总通信类别的百分比设定为：军事 30%、政治 30%、经济 30%和特殊 10%，按照通信内容对用户的影响程度从大到小设置为，特殊类＞政治类＞军事类＞经济类。各通信类型和内容类别的权重设置原则按照表 3 所示进行。实验中，为更好反映用户通信特点和模型对事件态势变化的把握能力，对各通信维度特征的重要性权重赋值均等，设定 Δt 为 1 min，对纵坐标做归一化处理，使用层次化 ICSI 态势评估模型进行相应的计算，得出事件级、区域级和系统级 3 个层次态势评估状况，分析如下：

(1)局域网仿真环境下的 ICSI 事件级态势(以 IP192.168.0.3 中的超长通话、频繁通信、通信类型和内容类别特征为例)

图 12 为 IP192.168.0.3 的超长通话、频繁通信、通信类型和内容类别特征在时间段 18：00-19：00 间的变化趋势。如图中所示，在 18：15 左右，内容类别特征达到接近 1 的态势值，而此时的频繁通信特征也保持在一个较高的水平，通信类型特征的归一化态势值则在零点附近。这说明此时该 IP 用户的通信过程中发生了具有较高权值的内容类型的事件，其采用的通信类型不具有较高的权值，且在此时段进行了多次通信，应引起网络管理员的重视，对应如图 13 所示，此时的事件级态势值为整个观测时段中的最高点。图 12 中，18：25 左右出现了内容类别特征的另一较高峰值点，但此时其他各维度特征归一化值均为零点附近。这说明此时发生的通信中出现了具有一定权值的内容类型的事件，但其不具有其他诸如多次通信等的特征，因此，可看作单次事件，对整个态势的影响力度不大，故此点的事件级态势值并不高，如图 13 所示。通过上述分析可知，本文模型和参数的计算方法结合了 ICSI 事件的多个维度特征，吻合用户和网络通信的实际状况，对 ICSI 事件的态势把握客观、清晰，便于网络管理员将关注点聚焦在那些尤为重要、影响度较大的事件上。同时，在实际应用中，可以根据网络信息安全管理的需求，对各个特征赋予不同的重要性权值，使其可以在多维特征的态势变化中得以重点显现，引起特别关注。

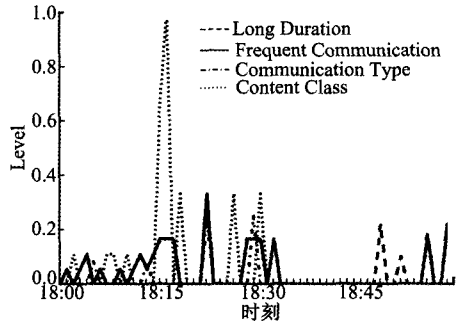

图 12　IP192.168.0.3 的超长通话、频繁通信、通信类型和内容类别特征态势

Fig. 12　Situation of lLong duration、frequent communication、communication type and content class characteristics of IP 192.168.0.3

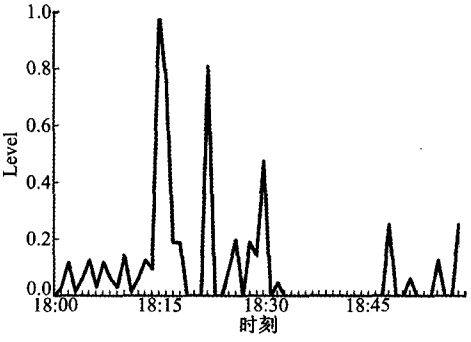

图 13　IP192.168.0.3 的事件级态势

Fig. 13　Incidents situation of IP192.168.0.3

(2)局域网仿真环境下的 ICSI 区域级态势

图 14 为 ICSI 的区域级态势。仿真环境下，建立了 3 个区域分别为 192.168.21、192.168.22 和 192.168.23，并分别赋予了不同的重要性权重。如图 14 所示，将区域内所属用户的多维特征综合，并根据赋予的各用户重要性权值形成的区域级态势可知，与实验一中的含有较少维度特征的态势相比较，实验二变化趋势的波动较多，图形

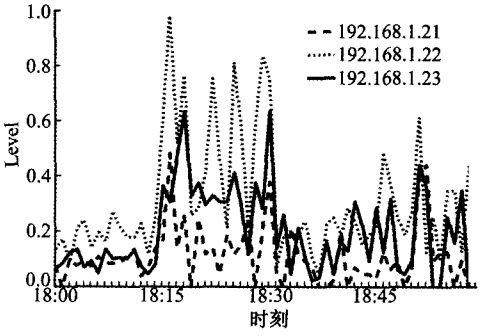

图 14　区域级态势

Fig. 14　Area situation

较为复杂。由此可以看出,ICSI事件本身具有多维特征,对其态势评估时结合的特征维度越多,越能更好、更完整地表达该事件的变化趋势和影响程度。

(3)局域网仿真环境下的ICSI系统级态势

图15为局域网仿真环境下的ICSI系统级态势图。从图中可以看出,局域网系统级态势的变化较实验一中两个开源数据集下的系统级态势变化曲线更为连续。这说明采用较小的时间窗口可以细粒度地反映网络中ICSI的变化,为信息安全管理提供有效的数据支持。

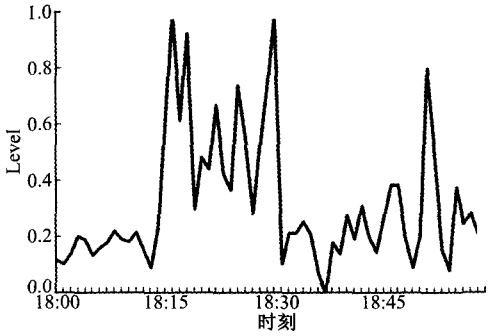

图15 系统级态势

Fig. 15 System situation

上述两个实验测试表明:

(1)本文提出的层次化ICSI态势评估模型具有有效性、可靠性和可行性。各个级别的态势评估结果与所涉及事件的多维特征及其重要性程度紧密相关,是一个全面、综合、系统的评估。

(2)对VAST 2008和Enron 2009数据集的测试结果说明,采用较大的统计分析时间窗口(以天为计量单位)可以提供较为宏观的事件态势评估走势图;同时,对于低维度数据集的态势评估的整体把握性强,可以从长时期的态势变化曲线中发现其中的安全规律。

(3)对局域网仿真数据集的测试结果说明:采用较小的统计分析时间窗口(以分钟为计量单位)可以提供较为微观的事件态势评估走势图;同时,对于高维度数据集的态势评估的特征敏感度强,能够从短时期的态势变化曲线中聚焦当前ICSI事件中的突出影响因素。

4 结束语

为解决网络信息内容安全事件的态势评估问题,本文提出了一种层次化的态势评估模型及参数计算方法。根据信息内容安全事件所具有的行为特征、内容特征、关系特征和位置特征,采用层次式结构模型。利用各特征内维度间的关系,对事件级、区域级和系统级态势评估值分别进行计算。为更好说明模型和方法的有效性和可行性,分别采用开源数据集VAST 2008、Enron 2009和局域网仿真环境下的数据集进行实验,结果表明:该模型和参数计算方法可以实现对网络信息内容安全事件的态势评估,体现了事件的强度和变化趋势,能够使网络管理员及时了解系统内的网络信息安全动态。下一步工作的重点是,将模型和方法应用于多网段局域网和大规模网络系统的信息内容安全事件的态势评估。

参考文献:

[1] GB/Z 20986—2007. 信息安全事件分类分级指南[S].

[2] 中国互联网络信息中心. 2013年中国网民信息安全状况研究报告[R]. 中国互联网络信息中心,中国,2013.

[3] 陈训逊,方滨兴,胡铭曾,等. 一个网络信息内容安全的新领域-网络信息渗透检测技术[J]. 通信学报,2004,25(7):185-191.
Chen Xun-xun, Fang Bin-xing, Hu Ming-zeng, et al. A new field in security of internet information and content-network information penetration detection technology[J]. Journal of China Institute of Communications, 2004, 25(7): 185-191.

[4] Fang Bin-xing, Guo Yun-chuan, Zhou Yuan. Information content security on the Internet: the control model and its evaluation[J]. Science China, 2010, 53(1): 30-49.

[5] 万源. 基于语义统计分析的网络舆情挖掘技术研究[D]. 武汉:武汉理工大学计算机学院,2012.
Wan Yuan. Research on mining of internet public opinion based on semantic and statistic analysis[D]. Wuhan: School of Computer Science and Technology, Wuhan University of Technology, 2012.

[6] Barroso N, Lopez de Ipina K L, Ezeiza A, et al. An ontology-driven semantic speech recognition system for security tasks[C]// Proceeding of IEEE International Carnahan Conference on Security Technology, Barcelona, 2011: 1-6.

[7] Roesch M. Snort-lightweight intrusion detection for networks[C]// Proceedings of the 13th System Administration Conference, Seattle, 1999: 229-238.

[8] Porras P A, Fong M W, Valdes A. A mission-impact-based approach to INFOSEC alarm correlation

[C]// Proceedings of the 5th International Symposium on Recent Advances in Intrusion Detection, Zurich, 2002: 95-114.

[9] Hariri S, Qu G Z, Dharmagadda T, et al. Impact analysis of faults and attacks in large-scale networks [J]. IEEE Security and Privacy, 2003, 1(5): 49-54.

[10] 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.
Chen Xiu-zhen, Zheng Qing-hua, Guan Xiao-hong, et al. Quantitative hierarchical threat evaluation model for network security[J]. Journal of Software, 2006, 17(4): 885- 897.

[11] Arnes A, Valeur F, Vigna G, et al. Using hidden markov models to evaluate the risk of intrusions[C] // Proceedings of the International Symposium on the Recent Advances in Intrusion Detection, Hamburg, 2006: 145-164.

[12] 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.
Wei Yong, Lian Yi-feng, Feng Deng-guo. A network security situational awareness model based on information fusion [J]. Journal of Computer Research and Development, 2009, 46(3): 353-362.

[13] Ahmed M S, Al-Shaer E, Taibah M, et al. Objective risk evaluation for automated security management[J]. Journal of Network and Systems Management, 2011, 19(3): 343-366.

[14] IEEE VAST 2008 Challenge[EB/OL]. [2008-03-15] https://www.cs.umd.edu/hcil/VASTchallenge08.

[15] Enron Email Dataset[EB/OL]. [2011-04-02] https://www.cs.cmu.edu/~enron/.