

# 基于本体模型的网络系统安全参数采集过程

李涛<sup>1</sup>, 林九川<sup>2</sup>, 胡爱群<sup>1</sup>

(1. 东南大学信息科学与工程学院, 江苏 南京 210096;  
2. 公安部第三研究所, 上海 200031)

**摘要:** 提出一种通用的网络系统安全参数采集过程, 将层次化的安全参数采集框架和本体化采集模型融入到整体网络架构中, 根据实际网络环境对安全数据进行属性分类和关联性描述, 对安全参数进行了多维度扩展, 抽象出信源、层级等多种属性, 有利于对参数进行进一步的分析。在此基础上, 描述了在实际系统中应用本框架进行参数采集的步骤, 能够有效地为整体网络安全评估提供指导和基础参数的支撑。

**关键词:** 安全评估; 安全参数采集; 本体论; 系统安全

**中图分类号:** TN929.53

**文献标识码:** A

**doi:** 10.11959/j.issn.2096-109x.2017.00131

## Ontology model based on security parameters capturing process for network systems

LI Tao<sup>1</sup>, LIN Jiu-chuan<sup>2</sup>, HU Ai-qun<sup>1</sup>

(1. School of Information Science and Engineering, Southeast University, Nanjing 210096, China;  
2. The Third Research Institute of Ministry of Public Security, Shanghai 200031, China)

**Abstract:** A general security data capturing process for network system was proposed, which combined hierarchical capturing framework and ontology model to whole network architecture. Attributes of security parameter were divided and relationships of parameter were described. Security parameters were extended to multiple dimensions. Multiple attributes such as information source and hierarchy were abstracted, which were benefit for analysis of security parameters. Based on the proposed framework, the process of application in real systems were described. The whole processes and framework efficiently provided guidance for network security evaluation and supporting of basic parameters.

**Key words:** security evaluation, security parameter capture, ontology theory, system security

### 1 引言

网络空间安全不仅关乎经济安全、社会稳定, 甚至关乎国家安全。掌握大规模网络系统的

安全状况, 保护网络系统正常运转至关重要。然而, 大规模网络系统的异构性、多样性和复杂性, 使网络空间的安全评估面临严峻挑战<sup>[1]</sup>, 主要原因在于网络安全参数的采集途径和手段单一, 缺

收稿日期: 2016-10-10; 修回日期: 2016-11-25。通信作者: 林九川, linjiuchuan@stars.org.cn

基金项目: 国家自然科学基金资助项目(No.61601113); 国家重点基础研究发展计划基金资助项目(No.2103CB338003); 公安部第三研究所开放课题基金资助项目(No.C15606)

**Foundation Items:** The National Natural Science Foundation of China (No.61601113), The National Key Basic Research Development Program(No.2013CB338003), The Open Research Fund from the Third Research Institute of Ministry of Public Security(No.C15606)

乏统一的采集方法和参数刻画标准,从而不能为整体安全评估提供有效的基础数据支撑,无法形成全局性、可量化性的安全度量标准。

网络系统中已有的漏洞扫描、入侵检测、流量分析、恶意代码检测、渗透测试等设备和工具已经能够提供安全参数采集的途径,但这些设备产生的安全相关参数处于分散、孤立的状态,缺少立体、全面的方法和理论模型。在安全参数采集方面,目前的研究主要针对单个安全属性参数,如 UC Berkeley 的 Ariel 和 Randy 针对网络设备的日志信息获取提出一种层次化处理方法<sup>[2]</sup>,有效地对大规模、分布式的安全日志进行获取和统一分析。

涉及多种安全参数采集和处理的研究主要是面向异常流量、僵尸网络和高级持续威胁检测,由于这些事件关系到多个网络数据的融合,因此需要对多种安全参数进行关联性分析<sup>[3~5]</sup>。国内有多家研究机构进行大规模网络安全相关参数获取和分析方法研究,如国家计算机网络应急技术处理协调中心(CNCERT/CC)以及北京大学的狩猎女神项目通过在可控安全实验环境中部署蜜罐系统进行僵尸程序检测<sup>[6]</sup>。现有一体化安全管理平台中涉及了多种网络安全参数的采集方法,但由于其在基础数据感知能力方面并没有明显优势,感知数据的精确性和全面性无法得到保证,在数据采集方面也缺乏有效的理论成果的支撑。

在安全参数的处理方面,传统方法实施的依据主要有两类:基于信息安全相关标准的符合与违反,基于网络系统关键部件控制权的争夺。从网络系统安全运维与保障的角度出发,可以将上述两类依据统一视为一种网络系统安全策略。目前,常见的信息安全相关标准(如 ISO27K 和等级保护系列)主要基于自然语言描述,要规范化这些安全要求语句以及它们之间的逻辑关系,保证这些规则可以被运用在任何网络系统安全评估过程中,则需要对其进行结构化,形成形式化的网络系统安全要求规约。

由上述分析可以看出,由于缺少统一理论框架的指导,安全参数的采集没有形成统一的体系,都是针对单一目标或几个目标对获取数据进行挖掘,其抽象和分析过程也是针对单一的数据进行

处理。安全属性基本以自然语言形式刻画,缺乏形式化的描述方法。安全特性的定义含糊、安全属性与安全特性之间以及安全特性彼此之间的关系描述不明确、缺乏清晰度、不同属性的安全参数之间无法进行统一,缺乏系统性的应用理论基础支撑。这些问题将损害基于安全参数采集进行整体网络安全评估过程的严谨性和客观性。

描述逻辑和 OWL 均是面向对象本体建模的常见语言之一,它们能够将特定对象的概念、属性以及约束关系进行形式化描述,构成清晰明确的规范说明<sup>[7,8]</sup>。本体(ontology)最早来源于哲学的概念,后来这个哲学范畴被人工智能界赋予了新的定义,从而被引入到信息科学中<sup>[9~11]</sup>,用于解决领域概念的形式化表述、概念属性之间的关系清晰刻画,以及复杂性特征问题的建模,并形成了统一的本体概念,“本体是共享概念模型的明确的形式化规范说明”<sup>[12,13]</sup>。安全本体概念主要是 Landwehr<sup>[14,15]</sup>提出的,后续的研究主要是对其提出安全本体的扩充。

本文在本体论的指导下,研究并提出面向整个网络系统的、多目标的、立体化的安全参数采集方法和机制,制定提供一个立体化的全面相应的采集、转换和呈现标准规范和相应的理论技术和方法,综合漏洞扫描、入侵检测、流量分析、恶意代码检测、渗透测试等安全参数采集途径,采用本体论的方法对安全参数抽象,有利于将来源不同的各种网络安全参数进行整理获取和融合。旨在解决目前大规模网络环境下安全参数分散、孤立的问题,形成统一的规范化参数。

## 2 网络系统安全参数采集过程

现有网络系统安全相关参数存在分散、孤立的问题,各大安全相关厂商已经形成固有的技术优势领域,因此需要对现有的网络安全设备进行功能扩展,研究数据规范化方法,以统一的标准对输出数据进行规范化改造,打破系统脆弱性评估方法采用的任务分发模式。在此基础上对网络安全相关参数进行融合处理,将数据分析的相关方法与网络安全对象和事件的本质特征相结合,增强基于数据分析的态势感知能力。

大规模网络包含多种网元要素,其部署方式

和产生的网络安全参数也存在差异性，因此，需要从整个网络系统的架构考虑，设计统一的采集框架，适用于各类网络系统安全信息的采集。从整个采集流程分析，对安全参数的采集包含以下步骤（如图1所示）。

1) 基于网络现有结构查看可以获取的网络安全相关参数，根据采集框架和评估目标研究是否需要增加新的采集设备、工具、软件等。

2) 对整体网络结构信息进行获取，包括网络的拓扑结构、服务器、无线接入点、活动端口、提供的服务、部署的安全相关设备和工具等。这些基本的结构信息和网元信息是产生本体属性的基础。

3) 根据整体采集框架抽象出安全需求，指导描述基本本体属性之间的相互关系，完成本体模型的构建。

4) 基于本体模型对获取的网络安全参数进行抽象描述，并进一步进行去冗余、相关性分析等操作，对数据进行融合呈现。

### 3 层次化安全参数采集框架

在采集架构的基础上，包含对安全相关参数的获取和分析的过程。数据获取是基于部署在网络中的安全相关工具、设备、软件等进行主动或被动的参数数据搜集。在全面获取数据的基础上对网络安全相关数据进行融合分析处理，并根据整体模型的需要进行标准化呈现，为安全性度量和评估提供基础数据层面支撑，这就需要根据整个安全参数处理的过程研究采集方法。本文提出一种结合安全相关数据感知、数据融合、数据呈现的一整套网络系统安全态势采集方法，从多维度、多层次、细粒度方面对网络安全参数进行本体化获取与处理。

在研究现有各类网络系统结构以及安全参数的特征和关联性的基础上，构建网络系统安全参数的采集框架，如图2所示，该框架包含3个层次：感知层、融合层和呈现层。

感知层对现有恶意代码检测、漏洞扫描、入

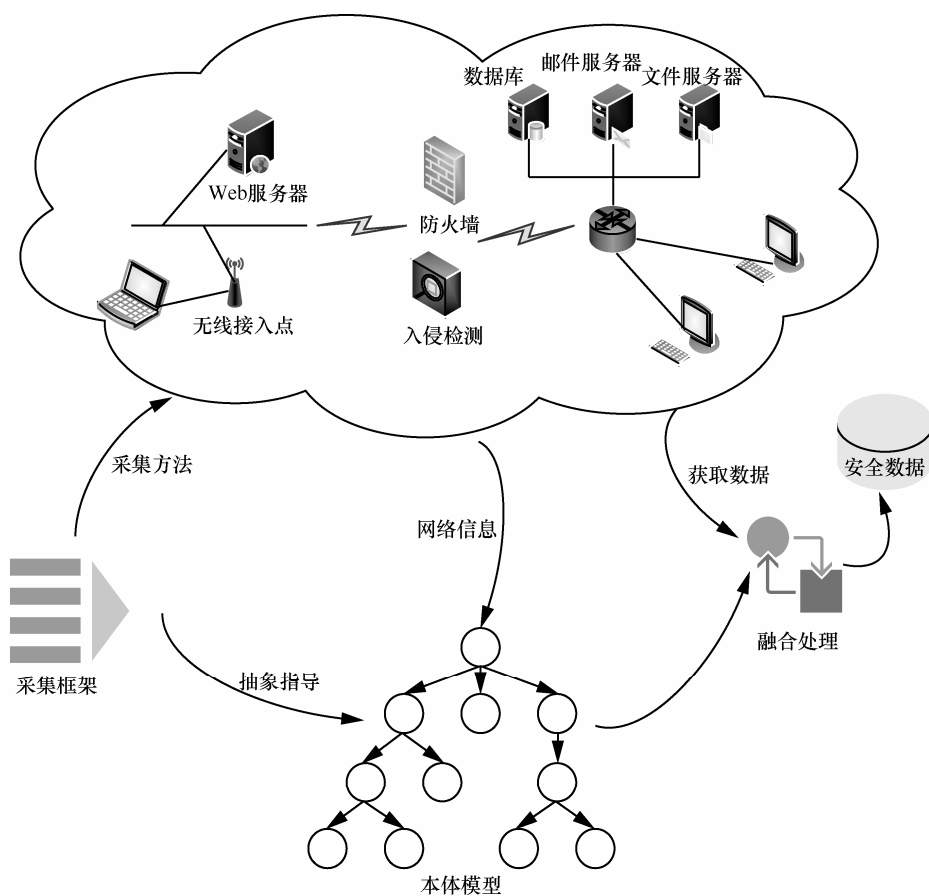


图1 基于本体化模型的安全参数采集架构

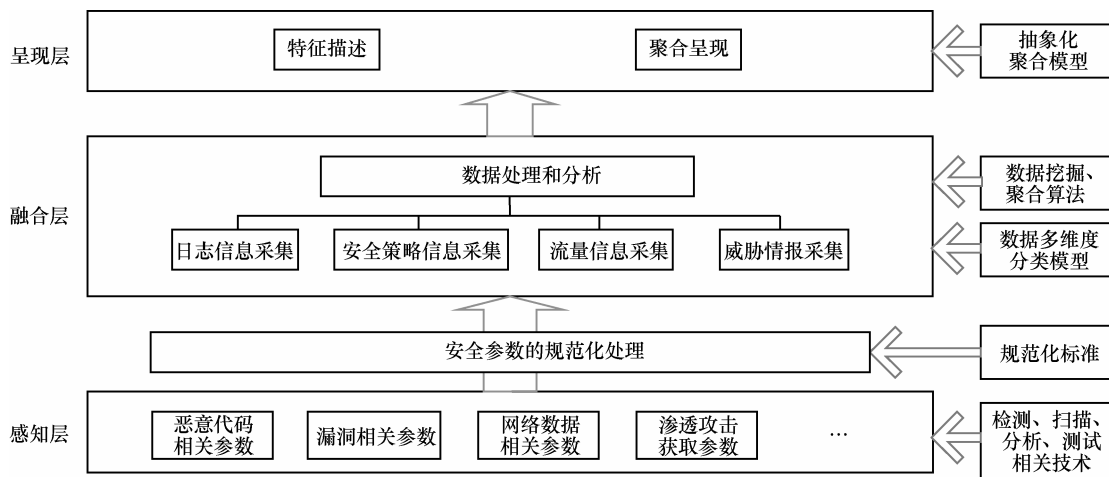


图2 安全参数采集框架

侵检测、渗透测试、网络异常行为分析等安全相关设备和工具进行扩展，在不破坏现有网络系统的基础上增强立体感知能力，获取不同粒度的网络指标参数，提升面向采集对象的漏洞自动检测及判别能力。并根据统一标准对数据输出接口进行规范化改造，对安全相关日志、参数、报告等进行规范化处理，为数据融合提供基础支撑。

融合层对多种安全参数进行分析处理，根据数据来源进行本体化描述，对网络系统的各种痕迹参数和安全事件的关联性进行分析，去除冗余，从数据表象发现本质安全问题和安全态势。为了提升分析效率和准确性，采用粗粒度和细粒度结合的机制，先对数据使用多维度采样分析进行定位，再使用深度分析进行数据挖掘，触发参数的精细化调整，必要时通过人工方法进行辅助决策。

呈现层将分析处理的数据按照整体度量体系的要求进行封装和聚合，为整体网络安全评估模型提供输入参数，根据现有标准中关于数据呈现的方法，结合整体评估体系和模型的需要，对聚合后的数据进行不同粒度的抽象，面向网络安全评估需求对安全参数进行不同粒度的封装和聚合，为评估模型提供输入参数和决策支持数据。

#### 4 信源安全参数本体化提取

本体在网络安全参数采集方面的作用主要有两点。一是利用本体提供准确的属性集，为数据采集提供引导，减少冗余和无用知识的产生，从而为参数处理更精准的知识发现。二是借助建立

完善的领域本体知识体系，对数据挖掘的维度进行扩展，从而得到更加全面、新颖的规则集。具体来说，可以结合具体的网络架构，分析与抽取其表述的常见形式与结构，研究描述网络安全评估领域内的基本概念，以及概念之间的关系，使用网络系统安全评估领域本体来形式化地描述网络安全参数采集的过程方法，以形成一种描述清晰、无二义性的网络系统安全参数分析规约，从而为网络安全参数采集建立领域本体模型并研究形式化的安全模型校验方法。

因此，在对获取的安全参数进行处理方面，首先需要研究现有相关检测标准中对安全参数的定义和规范化描述，找出各种安全相关数据的通用性和差异性，在此基础上制定出一种具有广泛适用性、简洁、充实的本体化模型。基于标准化处理后的安全参数，结合整体度量框架，利用本体抽象的方法表达各种参数之间以及参数与整体度量模型之间的关系，抽象出参数的内在属性和继承关系，对行为特征相同的数据、继承关系低于关联门限值的数据进行去冗余化处理，最后对有效数据进行聚合和呈现。

##### 4.1 安全参数采集的本体化模型

在安全参数采集的整体框架中构建了一种包含感知层、融合层和呈现层的安全参数处理过程，其中感知层获取的是直接的网络安全参数，数据种类复杂、多变，对其进行本体化建模尤为重要，本体化建模包含了本体基本要素的分类和要素之间依赖关系的描述。

根据大规模网络系统的构成元素, 本文描述了一种通用的网元本体化分类架构, 如图3所示。根据网络的基本要素抽象出安全要素分类, 将网络元素划分为4个区域: 主机域、网络域、数据域和应用域, 基于此, 对4个域的安全参数进行本体化抽象和关系描述。

主机域为分布在网络空间中的终端实体, 产生的安全相关参数包括终端已经部署的安全防护机制、检测到的恶意代码情况和检测到的漏洞、威胁情况。

网络域为从网络传输数据探测到的安全相关参数, 包括网络安全边界的安全域、防火墙相关的配置和日志数据、入侵检测设备的配置和探测到的威胁、通过主动渗透测试探测到的威胁、通过高级持续威胁检测探测到的攻击、检测到的异常流量数据、检测到的僵尸网络数据。

数据域为与应用数据的存储、传输相关的安全参数, 包括建立VPN通道的相关参数、数据库的安全相关参数、灾难备份系统的参数、文档加密安全的参数。

应用域为与应用相关的安全参数, 包括

Web安全检测到的威胁和漏洞、扫描到的垃圾邮件、进行域名安全检测和恶意网站检测获取到的参数。

上述网元安全参数分类是一个通用的架构, 可以方便地进行扩展, 根据具体网络安全环境和安全评估的需要进行子类的划分和层次的添加, 并基于本体化建模的需求, 进一步添加实例和依赖关系, 如对主机域中的漏洞、威胁检测进行本体关系扩展。如图4所示, 箭头表示的是类继承关系, 底层次要素“数据处理漏洞”分类到“源代码漏洞”中, 并属于一级要素“漏洞”的范畴。

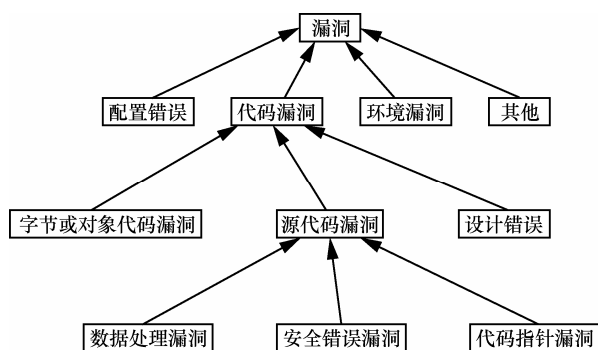


图4 漏洞要素的继承关系

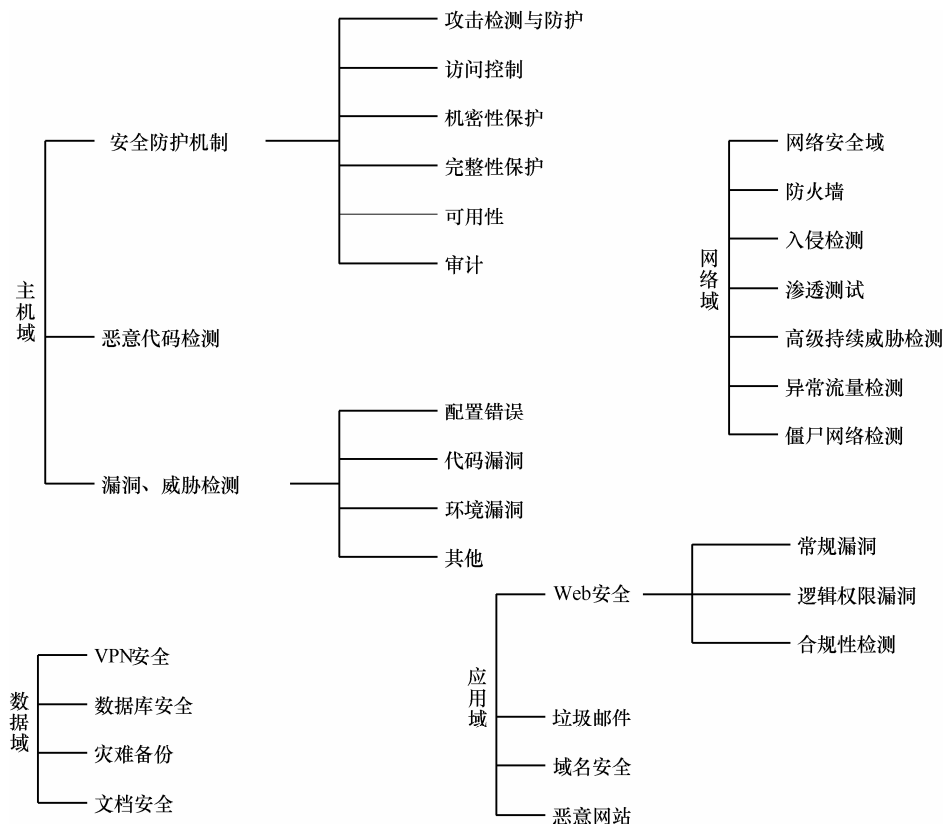


图3 网元安全参数分类框架

在对安全参数本体的基本要素和继承关系进行建模后,可以使用 OWL 语言<sup>[16]</sup>和 Protégé 本体编辑器<sup>[17]</sup>对对象本体和对应实例进行编辑,抽象成机器容易处理的语言,有利于进行自动化处理。

## 4.2 安全参数提取

在对安全参数的属性和关系进行本体化建模后,能够清晰地描述具体参数的分类和参数间的关系,在此基础上对参数进行提取处理。本体论本身是注释语法,提供常用的和直接的结构形式化地定义相关领域的属性资源<sup>[18]</sup>,在本模型中,对抽象的安全属性参数可以形式化地描述为一个五元组 $(C, H^C, R, H^R, I)$ ,其中,  $C$  代表一系列的基本安全属性或分类,这些属性根据  $H^C$  的定义进行层次化排列,  $R$  表示 2 个属性间的关系,  $R_i \in R$ , 并且  $R_i \rightarrow C \times C$ 。  $H^C$  和  $H^R$  分别描述了元素和关系的层次。  $I$  代表分类或关系的个体信息。

由于大规模网络环境下安全参数较为复杂,仅抽象出其本体化的属性和关系特征并不能满足后续数据处理的要求,需根据需要抽象其他特征属性。常用特征属性描述如下。

- 1)  $(C, H^C, R, H^R, I)$ , 本体化属性和关系特征。
- 2)  $S := \{D, T, C\}$ , 信息源参数, 其中  $D$  为日志信息,  $T$  为流量信息,  $C$  为配置信息。
- 3)  $EL := \{1, 2, 3, 4, 5\}$ , 安全参数所处的层级, 其中, 1 级为部署的设备、工具、软件等直接产生的安全异常事件, 如病毒扫描工具上报的病毒参数、漏洞扫描工具扫描的漏洞、入侵检测设备检测到的入侵事件等; 2 级为系统运行过程中的异常事件, 如主机异常关机、网络突然阻塞等; 3 级为通过简单分析发现的异常事件, 如通过数据分组地址的分析, 可以发现 DDoS 攻击来源和目的; 4 级为通过数据重组发现的异常事件, 如将网络传输的数据分组进行重组, 可以发现应用层存在的病毒; 5 级为通过深入分析发现的异常事件, 如通过关联性分析, 发现高级持续威胁的攻击。

4)  $CM := \{SecDev, NetDev, NetProbe, Software, ActiveDet, InfoGather\}$ , 获取安全参数的采集方法, 其中,  $SecDev$  表示从安全设备获取的参数, 如防火墙、入侵检测系统、漏洞扫描系统、病毒扫描工具等;  $NetDev$  表示从网络设备获取的参数, 如交换机、路由器、服务器等;  $NetProbe$

表示部署的网络探针获取的参数, 一般为主动部署, 用于有针对性地监控特定数据;  $Software$  表示部署在网络节点和终端的采集软件获取的参数;  $ActiveDet$  表示采用渗透测试这种主动式安全探测获取的参数;  $InfoGather$  表示采用信息汇聚进行综合分析后获取的参数。

通过上述抽象处理, 形成了安全参数的多维度、层次化的结构, 安全参数的来源、性质、所处的层级结构、处理的深度等信息都能得到体现, 为进一步的评估处理奠定了基础。

## 5 采集框架应用示例

上文提出了针对网络空间的安全参数采集的通过程, 并描述了其采集的框架和数据本体化处理的方法, 图 5 描述了将该框架应用到实际网络环境中的必要步骤。

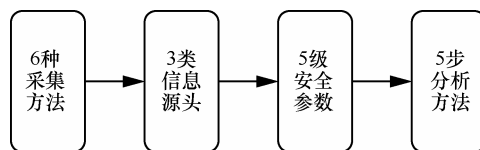


图5 网络安全参数采集分析步骤

首先, 根据整体网络安全评估的需要在网络中部署 6 大类采集方法, 形成安全设备、网络设备、网络探针、采集软件、主动安全探测和采集信息汇聚的多种采集能力。通过部署的采集方法获取安全参数后, 提取信息源包含的安全参数, 对 3 类信息源 (日志信息、流量信息、配置信息) 进行规范化处理, 按照处理的深度, 形成 5 级安全参数。

接着, 对数据进行分析, 分析过程如图 6 所示, 包含 5 个步骤: 1) 基于本体化理论模型对数据进行抽象, 刻画其多维度的特征; 2) 去除冗余操作, 由于网络安全设备数量众多, 很多数据都存在某种关联性, 若一个属性可以由其他相关属性推导出来或对整体评估没有帮助, 该属性就可以被认定为冗余属性; 3) 分区域、分维度进行分析, 横向上由点到面, 形成区域事件到整体事件的覆盖, 纵向上进行不同维度的分析, 建立不同维度的数据关联网; 4) 综合分析, 根据预处理的数据得到精炼的、较为直观的安全参数; 5) 将参数输入到整体安全评估模型中进行安全评估。

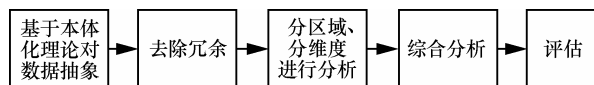


图6 网络安全参数分析步骤

## 6 结束语

本文提出了一种通用的网络安全参数采集过程，为安全参数采集过程构建层次化框架，有利于将现有的技术手段与全新的评估模型和方法进行结合。在数据规范化处理后以基于领域本体模型的网络安全评估方法为指导，将本体模型贯穿到安全数据的处理过程中。本体论的特点在于对对象特征的本质化描述，可以实现网络安全参数的多维属性特征提取，有利于清晰描述参数之间的关联性。这种全新的安全参数采集方法为安全参数获取提供了理论指导，有利于形成优化的、适用的安全评估方法。

在应用中，本框架实施的最大优势在于充分考虑了现有的安全工具，不同的安全厂商具有鲜明的技术特点，在病毒和漏洞扫描、防火墙和入侵检测、Web 安全检测等方面有各自的优势，在不改变现有设备功能的基础上对数据进行规范化处理，有利于为整个评估模型中安全相关参数的获取提供基础性支撑。利用现有的技术优势，通过进一步研究相关的规范化标准，对现有的数据进行规范化改造，能够解决现有网络系统安全相关参数分散、孤立的问题，有利于形成大规模网络环境下的立体化、全面的安全参数获取能力和安全态势感知能力。

## 参考文献：

- [1] 张焕国, 韩文报, 来学嘉, 等. 网络空间安全概述[J]. 中国科学: 信息科学, 2016, 46(2): 125-164.  
ZHANG H G, HAN W B, LAI X J, et al. Survey on cyberspace security[J]. Scientia Sinica Informations, 2016, 46(2): 125-164.
- [2] A system for reliable large-scale log collection[EB/OL]. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-25.html>.
- [3] JONGHOON K. PsyBoG: a scalable botnet detection method for large-scale DNS traffic[J]. Computer Networks, 2016, 97(14): 48-73.
- [4] KIRUBAVATHI G, ANITHA R. Botnet detection via mining of traffic flow characteristics[J]. Computers & Electrical Engineering, 2016, 50: 91-101.
- [5] ZHANG Y. An adaptive flow counting method for anomaly detection in SDN[C]//The 9th ACM Conference on Emerging Networking Experiments and Technologies. 2013: 25-30.

- [6] 王威, 方滨兴, 崔翔. 基于终端行为特征的 IRC 僵尸网络检测[J]. 计算机学报, 2009, 32(10): 1980-1988.  
WANG W, FANG B X, CUI X. IRC botnet detection based on host behavior[J]. Chinese Journal of Computers, 2009, 32(10): 1980-1988.
- [7] BHANDARI P, GUJRAL M S. Ontology based approach for perception of network security state[C]//Recent Advances in Engineering and Computational Sciences (RAECS). 2014: 1-6.
- [8] G'OMEZ -P'EREZ A, FERN'ANDEZ-L'OPEZ M, CORCHO O. Ontological engineering[M]//Metasynthetic Computing and Engineering of Complex System. Berlin: Springer, 2004.
- [9] GUARINO N. Formal ontology and information systems[C]//FOIS'98. 1998: 3-15.
- [10] FRANK A U. Spatial ontology: a geographical information point of view[M]. Berlin: Springer, 1997.
- [11] ARANGO G, WILLIAMS G, ISCOE N. Domain modeling for software engineering[C]//The 13th International Conference on Software Engineering. 1991: 340-343.
- [12] GRUBER T R. A translation approach to portable ontologies[J]. Knowledge Acquisition, 1993, 5(2): 199-220.
- [13] BORST W N. Construction of engineering ontologies for knowledge sharing and reuse[D]. Enschede: University Twente, 1997.
- [14] AVIZIENIS A, LAPRIE J C, RANDELL B, et al. Basic concepts and taxonomy of dependable and secure computing[J]. IEEE Trans. Dependable Sec. Computer, 2004, 1(1): 11-33.
- [15] LANDWEHR C E, BULL A R, MCDERMOTT J P, et al. A taxonomy of computer program security flaws[J]. ACM Computer Survey, 1994, 26(3): 211-254.
- [16] Owl Web ontology language[EB/OL]. <http://www.w3.org/TR/owl-features>.
- [17] The protege ontology editor and knowledge acquisition system [EB/OL]. <http://protege.stanford.edu>.
- [18] GRUBER T R. A translation approach to portable ontology specification[J]. International Journal Human-Computer Studies, 1993 (43): 907-928.

## 作者简介：



李涛（1984-），男，江苏镇江人，博士，东南大学讲师，主要研究方向为安全评估、移动终端防护。

林九川（1980-），男，江苏盐城人，公安部第三研究所副研究员，主要研究方向为系统安全评估、漏洞挖掘。

胡爱群（1964-），男，江苏如皋人，博士，东南大学教授、博士生导师，主要研究方向为网络与信息安全、物理层安全技术。