ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.21, No.7, July 2010, pp.1605−1619
doi: 10.3724/SP.J.1001.2010.03835
© by *Institute of Software, the Chinese Academy of Sciences*. All rights reserved.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

网络态势感知研究*

龚正虎, 卓 莹+

(国防科学技术大学 计算机学院,湖南 长沙 410073)

Research on Cyberspace Situational Awareness

GONG Zheng-Hu, ZHUO Ying+

(Institute of Computer, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: E-mail: zhuoying@nudt.edu.cn

Gong ZH, Zhuo Y. Research on cyberspace situational awareness. *Journal of Software*, 2010,21(7):1605−1619.
http://www.jos.org.cn/1000-9825/3835.htm

Abstract: The rapid development of Internet leads to an increase in system complexity and uncertainty. Traditional network management can not meet the requirement, and it shall evolve to fusion based Cyberspace Situational Awareness (CSA). Based on the analysis of function shortage and development requirement, this paper introduces CSA as well as its origin, conception, objective and characteristics. Firstly, a CSA research framework is proposed and the research history is investigated, based on which the main aspects and the existing issues of the research are analyzed. Meanwhile, assessment methods are divided into three categories: Mathematics model, knowledge reasoning and pattern recognition. Then, this paper discusses CSA from three aspects: Model, knowledge representation and assessment methods, and then goes into detail about main idea, assessment process, merits and shortcomings of novel methods. Many typical methods are compared. The current application research of CSA in the fields of security, transmission, survivable, system evaluation and so on is presented. Finally, this paper points the development directions of CSA and offers the conclusions from issue system, technical system and application system.

Key words: cyberspace situational awareness; data fusion; model; knowledge representation; assessment method

摘 要: 随着 Internet 规模的迅速扩大,复杂性和不确定性也随之增加,基于融合的网络态势感知必将成为网络管理的发展方向.在分析现有网络管理不足以及发展需求的基础上,介绍了网络态势感知的起源、概念、目标和特点.首先,提出了一个网络态势感知研究框架,介绍了研究历程,指出了研究重点以及存在的问题,并将现有评估方法分为 3 类:基于数学模型的方法、基于知识推理的方法、基于模式识别的方法.然后详细讨论了模型、知识表示和评估方法这 3 方面的研究内容,总结存在的共性问题,着重评价了每种评估方法的基本思路、评估过程和优缺点,并进行了对比分析.随后介绍了网络态势感知在安全、传输、生存性、系统评价等领域的应用研究.最后指出了网络态势感知的发展方向,并从问题体系、技术体系和应用体系 3 方面作了总结.

关键词: 网络态势感知;数据融合;模型;知识表示;评估方法

中图法分类号: TP393 **文献标识码**: A

* Supported by the National Basic Research Program of China under Grant No.2009CB320503 (国家重点基础研究发展计划(973)); the National High-Tech Research and Development Plan of China under Grant No.2008AA01A325 (国家高技术研究发展计划(863))

Received 2009-02-12; Revised 2009-07-06; Accepted 2010-03-04

© 中国科学院软件研究所 http://www.jos.org.cn

Internet 作为复杂巨系统,组网方式多样,传感器网络、AD Hoc、天基网等新型网络的加入,使得拓扑结构复杂化,难以准确获得;网络设备异构、数量巨大、移动性强;信息交互频繁,网络流量激增,网络负载增大;新应用不断涌现,VoIP,P2P,Grid 等应用的出现,构成了凌驾于传输网络之上的覆盖网络;网络时刻受到故障、攻击、灾难、突发事件的威胁,可用性、安全性和生存性面临严峻挑战;网络运行状况瞬息万变.随着 Internet 规模的迅速扩大,复杂性和不确定性也随之增加,对其特征做出有意义描述的能力相应降低.传统的网络管理各功能单元处于独立的工作状态,缺少有效的信息提取和信息融合机制,无法建立网络资源之间的联系,全局信息表现能力差.海量的网管信息非但不能加强管理,反而增加了网络管理员的负担.现代网络管理必须能够在急剧动态变化的复杂环境中,高效组织不确定的网管信息并进行分析评估,提供被管对象的详细信息,提高网络管理员对整个网络运行状况的认知和理解,提供多样化、个性化的管理服务,辅助指挥人员迅速、准确地做出决策,弥补当前网络管理的不足.

Bass 于 1999 年首次提出网络态势感知(cyberspace situational awareness,简称 CSA)的概念,并且指出,“基于融合的网络态势感知”必将成为网络管理的发展方向[1,2].所谓网络态势是指由各种网络设备运行状况、网络行为以及用户行为等因素所构成的整个网络的当前状态和变化趋势.值得注意的是,态势强调环境、动态性以及实体间的关系,是一种状态,一种趋势,一个整体和宏观的概念,任何单一的情况或状态都不能称其为态势.CSA 是指在大规模网络环境中,对能够引起网络态势发生变化的要素进行获取、理解、评估、显示以及对未来发展趋势的预测.

态势感知的概念源于军事需求,作为数据融合的一个组成部分——level 2 融合,是决策制定过程的重要环节.CSA 的目标是将态势感知的成熟理论和技术应用于网络管理,在急剧动态变化的复杂环境中,高效组织各种信息,将已有的表示网络局部特征的指标综合化,使其能够表示网络的宏观、整体状态,从而加强管理员对网络的理解能力,为高层指挥人员提供决策支持.与传统的网络管理系统相比,CSA 具有以下特点:(1) 运用数据融合技术,综合考虑影响网络态势的多种因素,提供全面而宏观的网络状态视图,加强对网络的理解与控制,以减轻网络管理员的负担;(2) 成为集成单元网管的平台,改变当前各单元网管独立工作的局面,实现信息共享,海量多样的信息带来丰富而准确的分析结果;(3) 为风险评估、决策制定提供支持.一旦完成态势感知,决策几乎可以根据态势自动生成.

本文系统而全面地介绍了网络态势感知的研究进展,主要贡献如下:(1) 提出 CSA 研究框架,概括了 CSA 的研究内容,总结了现有研究的特点和存在的问题;(2) 提出态势评估的分类方法,明确指出缺少统一规范的评价标准是目前评估方法复杂多样、各自为政的根源;(3) 系统讨论了 CSA 模型、知识表示、评估方法这 3 方面的研究重点以及应用的研究现状,指出了共性问题和现有研究的不足;(4) 对多种评估方法进行深入的分析和比较;(5) 指明未来研究趋势,从技术、功能和系统 3 个层面论述了“综合”是 CSA 的本质.

本文第 1 节提出一个 CSA 研究框架,指出 CSA 研究内容、现有研究的特点以及存在的问题,并且对各种评估方法加以分类.第 2 节介绍多种 CSA 模型以及模型的发展与演化.第 3 节讨论网络态势知识表示的两个问题:不确定信息表示和复杂系统表示.第 4 节在分类的基础上,详细介绍几种新颖的态势评估方法,并对各种典型的评估方法进行对比分析.第 5 节介绍 CSA 应用研究.第 6 节展望未来的研究趋势.最后从问题体系、技术体系和应用体系 3 方面对全文做出总结.

1 网络态势感知研究概述

本节在提出一个网络态势感知研究框架的基础上明确了 CSA 研究内容,并且提出了对现有评估方法进行分类的方法;然后概括了技术发展的历史进程;最后分析了现有研究的特点,并指出存在的问题.

1.1 研究框架

CSA 作为数据融合的一部分,并不是孤立存在的,向下从 level 1 融合获取各类网管数据,向上为 level 3 融合提供态势信息,用于威胁分析和决策支持,而且与其他融合层次关系紧密.层与层之间不仅数据通信频繁,而且方法相通,没有明确的界限,作为一个整体而存在.因此,CSA 研究包括多方面内容,其总体研究框架如图 1 所示.

© 中国科学院软件研究所 http://www.jos.org.cn

CSA 研究框架概括了 CSA 研究内容,体现了 closing-the-loop 理念,突出动态循环、不断细化的本质,强调反馈的重要作用.由图 1 可知,CSA 研究内容广泛,包括自身功能细化、关键技术的理论方法、与其他融合层次的通信与交互.目前的研究主要集中在 3 个方面:CSA 模型、知识表示和评估方法.其中:模型是研究的重点,相对比较成、熟统一;评估方法是 CSA 研究的核心,主要研究现有理论在态势评估中的运用;而有关知识表示的研究相对较少.根据不同的应用领域,网络态势又可分为安全态势、拓扑态势、传输态势、生存性态势等,从 Bass 开始,多数研究围绕安全态势展开.

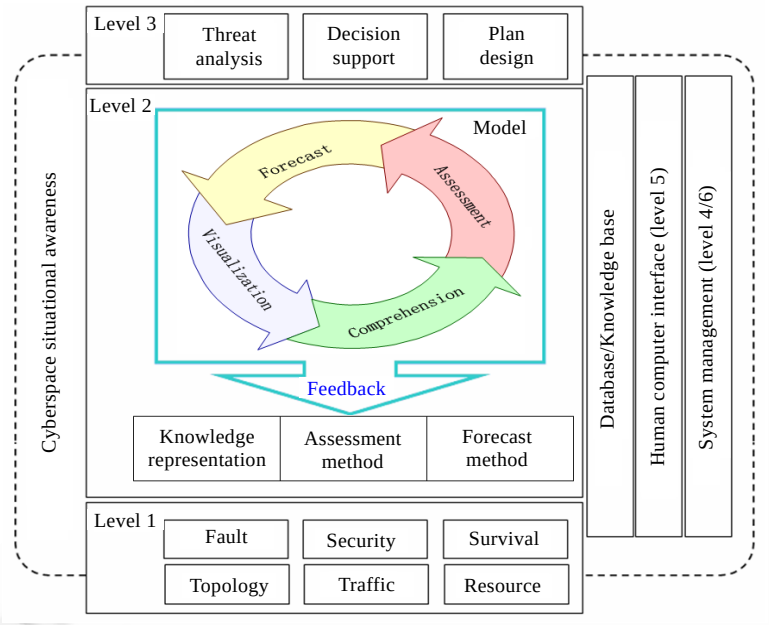

Fig.1 Research framework of cyberspace situational awareness

图 1 网络态势感知研究框架

1.2 评估方法分类

作为 CSA 研究的核心,评估方法多种多样,这些方法大致可以分为 3 类:

(1) 基于数学模型的方法

基于数学模型的方法(mathematics model,简称 MM)就是综合考虑各项态势因子,构造评定函数,建立态势因子集合 R 到态势空间 θ 的映射关系 $\theta=f(r_1,r_2,\ldots,r_n)$,其中,$r_i\in R(1\leqslant i\leqslant n)$ 为态势因子.态势包含众多相互冲突、不可公度、不确定的复杂态势因子,这些因子具有层次结构,可以逐层划分、细化.王娟[3]讨论了网络态势因子的组成与结构.传统、通用的多目标决策理论和效用理论的有关方法,如最大隶属原则、距离偏差法、打分法、多属性效用函数等,都可以用于态势评估.最常用的当属权重分析法以及公式法和集对分析.

MM 方法建立明晰的数学表达式,模型易于理解,而且能够建立连续的态势空间,给出一种有利或不利的判断性结果,便于态势的优劣对比.但是评定函数的构造、参数的选择没有统一科学的方法,一般依赖领域知识和专家经验,不可避免地带有主观意见,缺少科学客观的依据.此外,态势评估多数情况使用自然语言表述知识,而这种知识不容易被转化为易于机器处理的数学表达式.因此,建立面向自然语言条件陈述的数学模型也成为该方法的难点.

(2) 基于知识推理的方法

基于知识推理的方法(knowledge reasoning,简称 KR)的基本思路是,在已知经验知识、先验概率的前提下,接收 level 1 融合的输出,根据实时监测的数据信息,通过一定的关系逐级推理得到对当前态势的判断,可以对态

© 中国科学院软件研究所 http://www.jos.org.cn

势空间进行划分,给出分级或分类结果.KR 方法又可以分为基于产生式规则的逻辑推理、基于图模型的推理和基于证据理论的概率推理,代表性的方法有模糊推理、贝叶斯网络、马尔可夫过程、D-S 证据理论.

KR 方法能够模拟人类思维方式,较之 MM 方法,将知识的运用融入推理的过程之中,具有一定的智能,类似于专家解决问题的过程.评估的结果建立离散的态势空间,能够确定态势优劣等级,或者能够指明态势的类型,一目了然,便于态势的理解和把握.

该方法的难点在于如何获取建立模型所需的知识.如果凭借经验会带有强烈的主观性;如果通过机器学习,相关的研究还比较少,而且如何学习到“模拟人类思维方式”的知识更是难上加难.可见,该方法的优点反而成了最大的障碍.此外,该方法维护大量推理规则,空间开销和推理代价都很高,如何应对大规模的问题是另一个需要考虑的问题.

(3) 基于模式识别的方法

基于模式识别的方法(pattern recognition,简称 PR)分为建立模版和模式匹配两个阶段.第 1 阶段建立态势模版,在对态势空间进行划分的基础上,识别所有可能出现的态势状态.划分没有统一的标准,可以将态势分为不同类型,也可以对态势进行分级.第 2 阶段模式匹配,通过计算实测数据与模版数据之间的关联,如果两组数据符合,或者关联系数达到预先规定的阈值,则认为匹配成功,从而确定态势状态.建立模版是 PR 方法的重点,关键在于选择分类方法.除了凭借专家经验、领域知识以外,机器学习也是划分的主要手段,从训练样本或案例中获得有关分类的知识,代表性的方法有基于案例的推理[4]、神经网络、灰关联分析、粗集理论、聚类分析.这些方法一般也用于模式匹配.

PR 方法引入机器学习机制,科学、客观,可以方便地从历史数据或者案例中获得态势划分的知识.但是该方法计算量大,在非实时环境中有很好的效果,但是在实时环境中可能无法满足要求,某些研究采用启发式算法提高效率.此外,由于分类知识是从历史数据中通过机器学习获得,机器很难给出直观的解释,不利于理解.

1.3 研究历程

网络态势感知作为数据融合的一部分,其研究首先引入数据融合模型,对原有模型进行修改细化,建立 CSA 模型,明确 CSA 的功能.有关模型的研究并不多,在 2006 年之后,CSA 模型已经比较完善,趋于成熟,相关研究也没有显著的进展.评估作为 CSA 功能的核心,其研究浩如烟海.MM 方法最早被用于态势评估.MM 着眼于影响网络态势的不同因素,通过集成多种态势因子,从多角度反映态势的状态;但是 MM 只解决了多属性融合,却没有涉及多源数据的融合,而且 MM 采用的模型固定,只能得到确定的评估结果,忽略了不确定性因素.为了解决上述 MM 存在的两大问题,KR 方法随之出现.KR 一方面借助模糊集、概率论、证据理论等处理不确定性信息;另一方面通过推理汇聚多源多属性信息.在 2002 年~2005 年间,以贝叶斯网络为代表的 KR 方法成为研究的热点,涌现出大量文献,重复研究严重.如何获取推理规则、先验概率,尤其是对 CSA 这样一个新型的研究方向,成为摆在 KR 面前最大的挑战,于是 PR 方法应运而生.PR 从 2005 年逐渐受到关注,凭借其强大的学习能力,从训练样本或者历史数据中挖掘态势模式划分的知识,科学、客观.现有大多数评估方法或多或少都会引入数据挖掘的思想,并且展露出综合使用多种方法的趋势.评估方法的发展体现了“出现问题→求解问题”的探索过程.而有关知识表示的研究则寥若晨星,而且开始得较晚.

1.4 研究特点和存在的问题

网络管理需求和广阔的应用前景,共同奠定了 CSA 的重要地位,有关研究也不断深入.从已有 CSA 的初步探索中可以看到,相关研究呈现如下特点:(1) 网络安全态势感知是研究的热点,其他领域,如流量、故障、拓扑、生存性则很少涉及;(2) 系统架构是研究的重点,相对成熟,尽管存在差异,但基本上接受了 JDL(joint directors of laboratories)数据融合模型和 Endsley 态势感知模型的设计思想;(3) 对网络系统的表示,以层次化结构为主;对不确定信息的表示,多采用简单的分级,化为离散型数据;(4) 评估方法以权重分析为主,有些研究尝试把数据融合的数学方法引入 CSA.

与此同时我们也看到,CSA 的研究才刚刚开始,还存在很多问题:

© 中国科学院软件研究所 http://www.jos.org.cn

其一,现有研究过分关注安全态势,未能集成现有的各单元网络管理技术,无法实现对全局态势的综合评估与展现.由于缺乏对整个CSA系统做出全面而系统的研究,使得“态势”的概念太过狭隘,没有体现态势的整体性和宏观性的特点;

其二,知识表示不够健全.使用层次结构表示网络系统尽管简单直观、易于分析,但是无法展现网络元素之间错综复杂的关系,不利于挖掘多源多属性数据内部潜在的态势信息.此外,在信息表示方面,如何选择并且扩展用于态势评估的特征测度、建立合理完善的指标体系,有待研究;

其三,态势评估缺少统一的标准.这部分是因为态势的概念比较抽象,至于什么样的态势为好,好到什么程度,往往只是一种感觉,无论打分还是分级都缺少科学依据,说服力不强;然而究其更深层次的原因,既没有明确、形式化的态势定义,又缺少度量评估结果优劣的指标和方法,无法形成对态势以及态势评估的共识.没有标准必然造成评估方法多种多样、千差万别,数据融合领域几乎所有的理论方法都被应用到态势评估阶段,而且不断有新方法被引入.这其中包括大量重复研究,甚至为了增加理论深度而使用某种数学方法.只有统一了评价标准,评估方法的研究才有了明确的目标和方向;

其四,现有研究只关注态势本身,相对独立.无论纵向(level 1/3)还是横向(level 4/5/6)都缺少与其他融合层次的联系,没有融入数据融合体系之中.从图1研究框架可以清楚地看出,目前重点研究的3个方面都是level 2融合的关键技术,而没有涉及通信、交互以及系统管理等内容,使得各层研究相互脱离,向下很难直接利用level 1融合的结果作为态势感知特征测度的原始数据,向上也没有迎合level 3融合的需求,为威胁分析、决策制定提供支持也无从谈起.

综上所述,目前网络态势感知尚处于研究的初级阶段,虽然受到学术界的普遍关注,成为数据融合领域的热点,但是大多数研究只涉及CSA的某个方面,而且还停留在理论探索阶段.全面而深入的理论研究和应用的实际部署是当前和未来研究工作的重点,而其中最亟需的就是建立统一的评价体系.只有明确了CSA的研究目标,其他研究才能以此为基础展开有意义的工作.

2 CSA模型

本节介绍了几种最具影响力的CSA模型,展示了模型的发展和演化过程,最后总结了模型研究的共性问题.

态势感知以数据融合为中心,其模型的建立也以数据融合模型为基础.目前已经提出了几十种数据融合模型[5]:智能循环、JDL模型、Boyd控制循环、Endsley模型、瀑布模型、Dasarathy模型、Omnibus模型、扩展OODA(Observe,Orient,Decide and Act)模型,此外还有知觉推理模型[6],依赖联想存储和关联数据库,模仿人类思维方式.

其中最具影响力的当属JDL数据融合模型[7].JDL模型将融合分为4个层次:目标细化、态势细化、风险细化、过程细化,其中:态势感知作为较高层次的level 2融合,向下从level 1融合接收网络元素的监测数据,作为态势感知的信息来源;向上为level 3融合提供态势信息,用于威胁分析和决策支持.Blasch[8]对JDL模型加以发展,在4层融合的基础上提出了第5层——用户细化,level 5融合强调用户的作用,需要用户的知识和推理.同时,为低层提供反馈,优化融合过程.目前,JDL模型进一步发展为DFIG(data fusion information group)模型[9].DFIG提出了第6层——任务管理,区分信息融合和管理功能.

Boyd控制循环模型(OODA),描述了目的与活动的感知过程,并将感知循环过程分为观察、判断、决策、行动这4个阶段.其中:观察从物理域跨越到信息域,判断和决策属于认知域;而行动从信息域回到物理域,完成循环.前3个阶段类似于JDL模型,行动阶段通过考虑决策在真实世界中的影响来闭合循环.扩展OODA模型提供了处理多并发、潜在交互数据融合的机制.

Endsley模型[10]对态势感知阶段作了进一步细化,分为态势觉察、态势理解、态势预测3个层次.目前,Endsley模型日渐受到关注,Salerno在其基础上提出了通用的态势感知框架[11],并结合网络环境的实际特点,分析了CSA的实现.

Tadda等人在通用态势感知参考模型的基础上,结合网络应用具体问题提出了Cyber SA模型[12].Cyber SA

© 中国科学院软件研究所 http://www.jos.org.cn

模型包含日志、配置、任务、攻击、入侵尝试等网络元素,体现出网络的特点.

尽管态势感知模型数目繁多,但是从中可以看到几点共性:(1) 模型的重点放在态势感知的功能上.除 Dasarathy 模型关注任务之外,其他模型都是对数据融合、进而对态势感知的功能进行划分.不同模型的组成部分尽管名称不同,但很多功能是一致的;(2) 循环是态势感知的本质.模型中每个组成部分,没有明确的顺序,而是重复迭代进行;(3) 强调反馈的作用.多数模型最终形成闭环系统.

3 网络态势知识表示

本节从信息表示和系统表示两个方面讨论了网络态势知识表示,并且指出现有研究存在的不足.

知识表示[13]解决两个问题:其一是对不确定信息的表示.信息表示关注的焦点是不确定性.信息不确定性的表示绝非态势感知所特有,相关的研究有很多,而且作为基础研究,提出了从表示到融合完整通用的解决方案,其中常见的不确定性理论有概率论、模糊集、可能性理论和证据理论等,已经在态势感知领域得到应用.而且,不确定性本身的形式很多,在不同的领域中其分类和定义也不一样.从语义上讲,基本上与 Klir 和 Yuan[14]的分类相差无几,即把不确定性分为模糊性(fuzzy)和多义性(ambiguity),而多义性又可分为非特异性(nonspecificity)和冲突(conflict).对应这些类型的不确定性,不同的不确定性理论所能处理的不确定性的种类也不一样.模糊集是处理模糊性的理论,概率论只涉及到事件之间的冲突;可能性理论表示出事件的非特异性,而证据理论描述了非特异性和冲突[15].

其中,使用最多的当数 Zadeh 教授提出的模糊集,是处理人类认知不确定性和不精确性问题的数学方法,可以对不确定的语义信息进行处理,是描述人脑思维处理模糊信息的有力工具;不仅可以独立使用构建表示和推理系统,还可以和其他技术相结合作为通用的信息表示技术.模糊集在态势感知领域得到认可并广泛使用.此外,灰色理论是一种利用灰色信息的理论方法,采用白化函数对各种灰色信息进行白化,从而得到灰色信息的量化估计.

另一个问题是对复杂系统的表示,这也是CSA的挑战之一.为了改变当前各单元网管独立工作的状况,建立真正意义上的 CSA,在很大程度上依赖统一的系统表示方法.能够对网络做出准确、全面、详尽的描述,是进行态势感知的前提.另一方面,网络作为复杂巨系统,面对纷繁丰富的内容和错综复杂的关系,对其进行描述的能力很弱,而相关的研究又很少.

本体论是其中的重要方法.本体源于哲学概念,被引入计算机科学领域后,特指对共享概念模型所作的明确而规范的形式化说明.Husserl 于 1985 年提出了形式化本体论(formal ontology).本体强调领域中的本质概念,同时也强调这些本质概念之间的关联,能够将该领域中的各种概念及概念之间的关系显式化、形式化地表达出来,从而表达出概念中包含的语义,增强对复杂系统的表示能力.在形式化本体论的基础上,Grenon[16]将空间相关内容 SNAP(spatial items of interest)和时间相关内容 SPAN(temporal items of interest)作为既有区别又有内在联系的两个组成部分分别考虑.Eric[17]使用本体论建立态势视图,从时空两方面分析给出形式化结构,但是过于具体,只起到层次化结构的作用.事实上,在 CSA 中广泛使用且能够和评估方法相结合、真正起到系统描述作用的系统表示方法还要数传统的树状层次结构,如 Bass[1]借用 SNMP MIB 树状结构模型表示安全威胁分类,建立 TCP/IP 威胁分类框架;陈秀真[18]构造层次化网络系统安全威胁态势,从上到下分为网络系统、主机、服务和攻击/漏洞 4 个层次.另外,我国学者蔡文提出的可拓学(extension sets)[19],以系统中相互矛盾的事件为研究中心寻求事物的内在机制,通过建立物元模型,把不相容的问题转化为相容问题来求解.

从以上分析可以看出,目前有关网络态势知识表示的研究还存在很多不足:(1) 研究缺少延续性,只是单纯地关注系统表示,而没有应用到后继的评估之中.以本体论为基础的系统表示方法,或者停留在理论层面上,提出通用的方法,过于抽象,无法和评估工作相衔接;或者过于具体,最终陷于层次化结构,无法表现复杂的关系,失去使用本体论的意义.这是因为本体论和可拓学都是系统的理论方法,放之四海而皆准,但其理论体系庞大,使用复杂;由于不是针对态势感知问题提出来的,涉及的表示元素不能充分体现网络态势的特点;(2) 树状层次结构尽管简单明了,体现了复杂巨系统固有的层次性,能够简化分析,但却无法表现复杂的关系和内容,不能满足

© 中国科学院软件研究所 http://www.jos.org.cn

态势感知的需要;(3) 作为态势感知的关键问题,信息表示研究不足,尽管对不确定性信息的表示已经隐含在评估方法之中.

4 网络态势评估方法

网络态势评估是指在大规模网络环境中,在 level 1 融合获取各类网络监测数据并进行简单处理的基础上,根据领域知识和历史数据,借助某种数学工具或者数学模型,经过分析推理,对由各种网络资源、网络运行以及用户行为等诸多因素构成的整个网络的当前状态做出合理的解释.态势评估强调实体之间的关系信息,决定了态势因子的汇聚方式,给出一种有利或不利的判断性结果,简言之,就是从态势因子集合到态势空间的映射.所谓态势因子集合,是指能够引起网络态势发生变化的因素的集合,它是监测指标集合的子集.

评估方法是态势感知乃至数据融合领域的重点,因此备受关注,理论研究相对成熟[20].有些研究将理论创新引入态势评估领域,有些则对传统方法进行拓展,还有的将多种理论综合运用,目的就是对不确定性信息进行分析,提高态势评估的准确性,同时还要在时间开销、评估代价等因素之间进行折衷.在众多评估方法中,传统方法包括贝叶斯技术、基于知识的方法、人工神经网络、模糊逻辑技术,引入的新理论有集对分析、D-S 证据理论、粗集理论、灰关联分析、聚类分析等.这些方法大致可以分为 3 类:基于数学模型的方法、基于知识推理的方法和基于模式识别的方法.

4.1 基于数学模型的方法

MM 方法综合考虑各种因素进行态势评估,其目标是从不同的角度评价网络态势.本节介绍 3 种常用的 MM 评估方法.

4.1.1 公式法

公式法起源于传统态势评估.为了解释战争行为,利用数学工具建立战争模型给出对抗的敌我双方力量动态损耗的经典表达式以及方程中系数的求法,揭示了双方兵力效能彼此消涨的基本规律,最具代表性的是 Lanchester 战斗模型.与此类似,在 CSA 领域,同样通过数学公式建立模型,以此解释网络行为和状态.例如,Power 公式 $Power=\frac{Throughput^{a}}{Delay}(0<a<1)$ [21],讨论了吞吐率和延迟之间的关系,评价了资源分配策略的有效性.

4.1.2 权重分析法

权重分析法是最常用的评估方法,其评定函数通常为指数表达式,由态势因子及其重要性权值共同确定.陈秀真[18]采用自下而上、先局部后整体的策略建立层次化网络安全威胁态势量化评估模型,在报警发生频率、报警严重性及其网络带宽耗用率的统计基础上,采用逐层汇聚的方式对攻击、服务、主机以及整个网络的重要性权值进行加权,计算威胁指数,进而评估安全威胁态势.

权重分析法是最典型的 MM 方法,兼有其优点和缺点,关键是求得态势因子的重要性权值.其最突出的优点就是将 level 1 融合的结果直接作为态势评定函数的参数,拉近了数据融合层次之间的距离.

4.1.3 集对分析

集对分析[22](set pair analysis)是我国学者赵克勤于 1989 年提出的一种关于确定不确定系统同异反定量分析的系统分析方法,所谓集对 H 是指具有一定联系的两个集合所组成的对子.联系数是集对分析的重要概念,其一般形式为 $U=A+Bi+Cj$,A,B,C 分别是关于所研究对象的同一性、差异性、对立性测度,联系数把 3 种测度联系在一起组成一个同异反系统(即确定不确定系统).除联系数以外,集对分析还定义了集对的势 $shi(H)=a/c$.集对分析可用于态势评估,构造量化分析模型,因为联系数中各个联系分量包含了系统的态势信息.

使用集对分析进行态势评估,其基本思想如下:首先确定系统的联系数 U 的表达式,给出同异反联系度 A,B,C 的计算方法,建立态势评估的集对分析模型;接下来分析 $shi(H)$的取值,可以判断系统在同异反联系中是否存在统一、对立或者势均力敌的趋势,进一步分析 A,B,C 的大小关系,根据排列组合原理,建立基于 $U=A+Bi+Cj$ 的系统态势状态表,A,B,C 三维系统态势有 12 种状态;最后根据集对分析模型计算系统在特定环境下的联系数,通过查找状态表确定系统态势所处的状态.

© 中国科学院软件研究所 http://www.jos.org.cn

联系数形式灵活,可以根据需要方便地加以展开.在系统态势评估中,如果对论域作同、异、反划分还觉得粗糙,那么可以将联系数展成 $U=A+Bi_1+Ci_2+Dj$ 的形式,进行四维态势分析,提供更为细致的 49 种可能状态.通过对 B 和 C 进行展开,用于论域的边界和内部都可作有限和无限细分的场合.当对问题的求解精度要求较高时,还可以引入 i 的取值分析,以考察那些细分界限的不确定性、模糊性是否会影响到结论的稳定性与可靠性.

集对分析的优点在于使用联系数统一处理模糊、随机、中介和信息不完全所致的多种不确定性;从同一性、差异性、对立性测度等多个角度进行态势评估,避免采用单一标准的局限性;而且基于联系数的系统态势分析是一种全排序,具有唯一性,对态势所处的状态级别进行明确划分,替代"取大取小"模糊评价方法,避免丢失大量有价值的信息、引发错误结论.尽管集对分析具有以上与生俱来的优势,但是作为一种 MM 方法,仍旧不可避免其固有的缺点,如何构造同异反联系度,始终缺少科学的依据和公认的方法,也因此成为集对分析的难点.

4.2 基于知识推理的方法

KR 方法充分利用经验知识建立态势评估模型,借鉴模糊集、概率论、证据理论等处理不确定性信息,通过逻辑推理判断网络态势完成评估.其目标是处理多源多属性信息.KR 又可以进一步细分,本节介绍几种经典的 KR 方法.

4.2.1 基于产生式规则的逻辑推理

基于产生式规则的逻辑推理是人工智能、专家系统的重要方法,然而经典的推理方法不能解决态势评估领域的不确定性,因此需要在逻辑推理中引入对不确定信息的表示方法.Zedeh 教授提出模糊集,正是处理人类认知不确定性和不精确性问题的数学方法,一般分为模糊化、模糊逻辑推理和去模糊化 3 个步骤,这是一种使用逻辑表达式来描述模糊集中隶属关系的推理方法,这里不再具体介绍.有关模糊逻辑推理的研究不断深入,连同其不同拓展,例如直觉模糊集、L-模糊集、区间值模糊集、Vague 集等,也逐渐被引入态势评估领域.

4.2.2 基于图模型的推理

基于图模型的推理,通过建立有向图表示状态转换.代表性的方法有贝叶斯网络和马尔可夫过程,前者用节点表示前提条件或结论,边表示节点间的逻辑关系;后者用节点表示可能的态势划分,边表示态势转换的条件.图模型中隐含了有关概率、关系、推理方法等知识.

贝叶斯网络也许是态势评估研究中使用最多的一种方法,无需再作介绍.相关研究仍不断展开,迎合态势评估动态、不确定性的特点,引入动态贝叶斯网络(dynamic Bayesian network,简称 DBN),进行关于时间的概率推理[23,24].动态贝叶斯网络表示时序模型,可以看成是一个隐马尔可夫过程,其区别在于,DBN 将复杂系统的状态分解成一些组成变量.DBN 能够充分利用时序模型中的稀疏性,降低空间和推理代价,适合大规模问题.同时,借鉴其他方法的思想,相继提出加权贝叶斯推理、层次贝叶斯推理[25].

基于图模型的推理突出了推理方法的优点,使用状态转换图表示推理过程,清晰明确、易于理解;但也扩大了推理方法的难点,图的存储开销大,如何通过机器学习建立图模型成为该方法亟需关注的问题.与之前介绍的方法相比,图模型建立在经典概率理论基础上,能够使用置信度表示不确定性,隐式解决了不确定性表示问题.

4.2.3 基于证据理论的概率推理

由 Dempster 和 Shafer 建立的证据理论[26],简称为 D-S 理论,是进行不确定性推理的重要方法.D-S 理论处理不确定性和无知性之间的区别,它并不计算一个命题的概率,而是计算证据可能支持命题的概率,给出信息的信任测度和似然测度,具有一定程度的怀疑能力.D-S 理论给出证据(即基本概率分配 BPA)和焦元(focal element)的定义,引入信任函数 $Bel(A)$表示证据给予命题 A 的全部信任程度,似然函数 $Pl(A)$表示证据不怀疑命题 A 的程度.其核心是 D-S 证据合成规则,给出方法将来自两个独立信息源的证据组合为一个新的证据,并且可以推广到多个证据的情况.

徐晓辉[27]将 D-S 理论引入计算机领域进行网络态势评估,其分析过程如下:首先建立证据和命题之间的逻辑关系,即实体、态势因子到态势状态的汇聚方式,确定基本概率分配;然后根据到来的证据,即每一则事件发生的上报信息,使用证据合成规则进行证据合成,得到新的基本概率分配,并把合成后的结果送到决策逻辑进行判断,将具有最大置信度的命题作为备选命题.当不断有事件发生时,这个过程便得以继续,直到备选命题的置信

© 中国科学院软件研究所 http://www.jos.org.cn

度超过一定的阈值,证据达到要求,即认为该命题成立,态势呈现某种状态.

使用 D-S 理论进行态势评估,克服了用概率描述不确定性的不足,不需要精确了解概率分布,也不需要显式表示不确定性.通过建立命题和集合之间的对应关系,把命题的不确定性问题转化为集合的不确定性问题,给出信息的信任测度和似然测度.当先验概率很难获得时,D-S 理论较概率论更为有效.使用 D-S 理论的另一大优点就是形式灵活多变,相关研究[28,29]将 D-S 理论和模糊逻辑、神经网络、专家系统相结合,进一步提高推理的准确性.该方法的缺点是计算复杂度高,而且实际应用中由于标准化而忽略了矛盾冲突,丢失了冲突信息,不适用于高冲突证据的情况.后者也成为 D-S 理论研究的热点,针对该缺陷提出了各种改进措施.

4.3 基于模式识别的方法

PR 方法通过机器学习建立态势模版,经过模式匹配,完成对态势的划分.其目标是不过分依赖专家和经验,自动获取知识,建立科学、客观的评估模版.本节介绍 3 种不同的 PR 方法.

4.3.1 灰关联分析

灰色系统理论[30]由我国学者邓聚龙于 20 世纪 80 年代建立,是一种处理不确定信息的理论方法.灰关联分析的基本思想是,根据序列曲线几何形状的相似程度来判断其联系是否紧密,曲线越接近,相应序列之间的关联度就越大,反之就越小.

将灰关联分析应用到态势评估领域,根据关联程度对态势状态进行模式关联和模式匹配.对态势进行灰关联分析,首先构造态势要素数据序列,选择模版序列和比较序列;然后计算每个因子的绝对差;进而计算两组态势序列的最大差和最小差,根据灰关联分析模型建立态势因子的灰关联系数,比较两组序列各个态势因子的关联系数值.但是,灰关联系数结果较多,信息过于分散,不便于比较.因此,接下来计算灰关联度,将每个序列的各个态势因子的灰关联系数集中体现在一个值上;灰关联度反映态势之间的关联程度,根据取值的大小对态势进行灰类划分.灰关联度的计算,可以选择各态势因子关联系数的代数平均值,也可以选择加权平均值,即根据态势因子的重要性为各因子分配不同的权重.

态势灰关联分析思路简单,提出了一种模式匹配方案.在各个态势历史数据之间进行比较,即可完成对态势的分类评估.但是在模式匹配阶段,需要与每一个模版进行比较,且每一次比较都需要计算灰关联系数和灰关联度,计算复杂性高.此外,没有涉及如何建立模版序列的问题.

4.3.2 粗集理论

粗集理论(rough set)由波兰数学家 Pawlak 于 1982 年提出[31],模拟人类抽象逻辑思维功能,是一种新型的处理模糊、不确定和不完备信息的数学工具.不同于传统处理方法,粗集理论建立在由等价关系对全域进行划分的基础上,研究表达知识系统的属性的重要性、属性之间的依赖关系、优选系统的描述特征,通过引入上近似集和下近似集来描述一个集合.其主要思想是,在保持分类能力不变的前提下,通过对知识的约简,导出概念的分类规则.

在态势评估中使用粗集理论,借助决策信息系统,通过决策表建立态势因子和态势划分之间的联系.在建立态势模版阶段,收集历史数据作为训练样本,经过特征选择、信息表离散化、属性约简、属性值约简、形式化规则提取 5 个步骤构造态势评估决策表[32];在模式匹配阶段,当新信息到达时,通过分类决策表可以确定当前态势状态[33].

基于粗集理论的态势评估,兼具表达、学习与分类能力,突出的特点在于粗集学习能力强,具有从海量历史数据或者案例中发现隐含知识、揭示潜在规律并转化为逻辑规则的优势.其次,借助信息系统这一形式化模型,将知识的表达、学习和分析纳入统一的框架之中.而且无须提供所需处理数据集合之外的任何先验信息,科学、客观,避免了主观因素带来的影响.其难点在于决策表核的确定和属性约简算法(求核与约简),属性约简就是在保持分类能力不变的条件下,删除其中不相关或不重要的知识.现已证明,求决策表所有约简和最小约简是一个典型的 NP 难题,计算量大,在非实时环境中有很好的效果,但在实时环境中可能无法满足要求,成为粗集理论研究的焦点问题.随着研究的深入和理论突破,基于粗集理论的态势评估方法也将不断完善,这一兼具表达、学习与分类能力的方法,必将拥有光明的发展前景.

© 中国科学院软件研究所 http://www.jos.org.cn

4.3.3 聚类分析

聚类是指根据数据的不同特征自动地将其划分为不同的簇(cluster),目标是属于同一个簇中的对象之间具有较高的相似度,而不同簇中的对象差别(相异度)较大.区别于分类是预先定义好分类的类别,聚类的类别取决于数据本身.聚类技术大致分为 5 类:划分方法(partitioning method)、层次方法(hierarchical method)、基于密度的方法(density-based method)、基于网格的方法(grid-based method)、基于模型的方法(model-based method).典型的聚类过程主要包括数据准备、特征提取和特征选择、相似度计算、聚类、聚类结果有效性评估等步骤.目前已经广泛应用于网络数据分析领域,例如异常检测[34]、关键流量矩阵发现[35]、使用模式挖掘[36]、文本搜索[37]以及态势评估[38]等.

聚类属于无监督机器学习方法,提供了一种对未知数据进行模式划分的方法,不需要任何先验知识,能够自动识别数据内在结构、发现隐藏在数据中的规律,科学、客观.尽管相关研究有很多,但是没有任何一种聚类算法可以普遍适用于揭示各种多维数据集所呈现出来的多种多样的结构[39].寻找适合网络数据流的聚类算法,是今后工作的重点.

4.4 典型方法比较

以上介绍了几种新颖的态势评估方法,综合其他传统的方法,从不同角度对它们进行对比,结果见表 1(因表宽所限,表中用 m.代表 medium),比较选择了 10 种指标,其中:建模时间、模型空间开销、评估时间反映了评估方法的效率以及动态更新模型的代价;特征数量、通用性、可扩展性反映了评估方法的灵活性,从侧面评价是否利于实际应用部署;根据结果形式评估方法可以分为打分、分级、分类 3 种;知识来源反映了评估方法是否客观,能够评价对经验知识和领域专家的依赖性;易理解性反映评估方法是否友好;并指出了理论固有的弊端.

Table 1 Comparison of different situation assessment methods

表 1 态势评估方法之间的比较

		Result form	Modeling time	Model space	Assessment time	Knowledge source	Number of feature	General	Scala-bility	Comprehen-sibility	Theory shortcoming
Maths model	Formula	Score	Short	Small	Short	Experience	Small	Low	Hard	Easy	Subjective
	Weighed	Score	Short	Small	Short	Experience	Small	Low	Hard	Easy	Subjective
	Pair set	Grade	Short	Small	Short	Experience+ Identical-Different-Contrary	m.	m.	Easy	m.	No uniform method
Knowledge reasoning	Fuzzy	Class	m.	m.	m.	Expert	m.	m.	m.	Easy	Subjective
	Markov	Class	m.	Big	m.	Expert + Probability	m.	m.	m.	Easy	Probability hypothesis
	Bayesian	Class	m.	Big	m.	Expert+ Probability	m.	m.	m.	Easy	Probability hypothesis
	Evidence	Class	m.	m.	Long	Probability+ Evidence combination	m.	m.	m.	Easy	Lose conflict information
Pattern recognition	Gray	Class	m.	m.	Long	Experience+ Gray relation degree	Big	m.	m.	m.	Subjective
	Neural	Class	Long	m.	m.	Machine learning	Big	High	Easy	Hard	Knowledge representation
	Rough set	Class	Long	m.	m.	Machine learning	Big	High	Easy	Hard	NP-Hard
	Clustering	Partition	Long	m.	m.	Unsupervised learning	Big	High	Easy	Hard	Low-Accurate

5 CSA 应用研究

CSA 从理论研究走向实际应用的探索才刚刚开始,从 Bass 首次提出态势感知的概念以来,相关研究绝大多数围绕安全态势展开,也有少量文献涉及流量分析、信息优势度量.近来,随着网络生存性日益受到关注,出现了对网络生存性态势感知的研究.目前,CSA 应用研究围绕态势感知核心部分展开,同样关注模型、知识表示、评

© 中国科学院软件研究所 http://www.jos.org.cn

估方法 3 个重点内容.安全领域研究相对深入,理论方法在评估中得到应用,有些方法实现了态势预测.尽管如此,层次结构与权重分析相结合仍旧是评估方法的主流.其他领域还停留在数据层面上,例如传输态势把重点放在检测工具开发和可视化研究上.而生存性态势还处于定性评估阶段,没有实现从数据到信息再到知识的抽象,也无法体现在网络管理中引入态势感知的优势和必要性.唯有 Blasch 关于 CSA 系统测试和评价的研究为 CSA 应用带来了新气象.作为 CSA 系统评价标准研究的具体案例,虽然只是给出一套特定的评价测度和计算公式,没有提出抽象通用的方法,也没有明确态势评估的标准,但却提出了一个新的课题方向,是 CSA 走向成熟规范的开始.总之,目前的研究只是 CSA 实际应用的初步尝试,相关工作尚未全面展开,有关全网态势感知、CSA 系统以至融合系统设计、态势与系统评价体系等课题仍有待研究.

5.1 Tim Bass的研究

Bass 有关 CSA 的研究[1,2]重点在于安全态势.在模型方面,借鉴 ATC(air traffic control)态势感知的成熟理论和技术,提出了基于多传感器数据融合入侵检测系统(IDS)的层次结构,分别给出了 IDS 数据融合模型和 IDS 数据挖掘模型,并且指出 IDS 必须采用 out-of-band 模式.在知识表示方面,Bass 区分了陈述式的知识和过程式的知识,前者是被动的知识或关系知识,如文件;后者是陈述式知识的一种特例,以模式、算法和数学变换来表示;一般陈述式知识基的容量远大于过程式的知识基.Bass 还借用 SNMP MIB 模型表示安全威胁分类,建立 TCP/IP 威胁分类框架.在评估方面,Bass 提出预防、检测、矫正的信息保障(information assurance,简称 IA)机制,建立基于重要性、脆弱性和威胁 3 大风险要素的风险识别模型.通过重要性和脆弱性矩阵对网络风险量化分析,推理识别攻击者身份、攻击速度、威胁性和攻击目标,进而评估网络的安全状态,并最终给出 9 步骤的统一风险管理过程,但没有实现具体原型系统[40].

5.2 安全态势感知

自 Bass 提出 CSA 概念之后,相关研究逐步展开,目前大多集中于网络安全态势感知.实现入侵行为检测、入侵率计算、入侵者身份和入侵者行为识别、态势评估以及威胁评估等功能,已经成为信息安全领域的一个新方向.安全态势关注网络的机密性、可用性、完整性,应用数据融合技术处理包括防火墙日志、入侵检测日志、病毒日志、网络扫描、非法外联、设备状态、实时报警在内的异类、多源数据.网络安全态势感知分为静态评估以及紧密结合环境、动态把握风险的动态评估.按照数据源,网络安全态势感知又可分为基于系统配置信息和基于系统运行信息两大类[18].陈秀真[18]使用层次结构建立威胁评估模型,结合权重分析法,逐层计算威胁指数并且绘制态势走势图,实现安全态势的量化评估及可视化.韦勇[41]利用改进的 D-S 证据理论将多数据源信息进行融合,使用权重分析逐层汇聚态势要素和节点态势计算网络安全态势,进一步结合实际性能信息修正节点安全态势值[42].此外,基于时间序列分析实现了网络安全趋势的预测.

5.3 传输态势感知

流量历来是网络管理的重点[38],在 CSA 概念提出之前,其研究已经相当丰富而且成熟,尤其是在流量测量与分析方面.从态势感知的角度,传输态势关注信息优势、可视化以及性能评价.信息优势[43]是指收集、处理和传播连续信息流的能力,同时削弱或剥夺对方收集、处理和传播连续信息流的能力,关注信息的完备性、准确性、时效性,评估方法多采用公式法和权重分析法.现有的 Internet 级网络流量可视化工具[44],提供整个网络的流量信息,易于发现网络攻击行为,提取攻击行为特征.其中,美国劳伦斯伯克利国家实验室的 Lau 于 2003 年开发了“The Spinning Cube of Potential Doom”[45]系统,该系统在三维空间中用点来表示网络流量信息:X 轴代表网络地址,Y 轴代表所有可能的源 IP,Z 轴代表端口号,提高了传输态势感知的能力.但是,可视化以原始流量数据为对象而不是高层抽象的态势信息,仍停留在数据层面.国内一些关于网络传输性能评价的研究和传输态势感知关系比较紧密,其中,杨雅辉建立了网络性能指标体系并给出形式化描述框架[46],江勇、林闯等人采用公式法以评价函数作为标准展开网络传输控制的性能评价[21,47],张冬艳等人以权重分析为基础评价网络性能[48,49].

© 中国科学院软件研究所 http://www.jos.org.cn

5.4 生存性态势感知

生存性是指在遭受攻击、故障或意外事故时,系统能够及时完成其关键任务的能力,与故障管理一起从不同的角度讨论网络的可用性,范围不局限于网络设施本身.在 9.11 事件之后,网络生存性受到广泛关注.生存性态势感知是对网络生存性能力的分析评估,关注完整性、可用性、保密性 3 个主要因素,尚未取得突破性进展.

5.5 CSA系统性能评价

不同于上述 CSA 研究关注网络管理的不同领域,Blasch[50]站在更高的角度研究态势感知系统的测试和评价方法,其目标是纯粹地从数据可视化或者负载减少的角度量化 CSA 的收益,测量系统的性能或者效率,判断是否满足任务需求.Blasch 通过引入测度的概念建立标准化的融合系统性能评价方法,提出了最小测度集合,并且明确了计算公式.Salernoa[51]将评价方法应用于 CSA,引入可靠性、纯度、代价、时间四维测度,建立评价函数,并且通过具体案例分析评价 CSA 系统的性能.

6 未来研究趋势

在详细讨论了 CSA 研究内容和应用进展的基础上我们发现,“综合”是 CSA 的本质,也决定了 CSA 的发展趋势.数据融合的提出就是为了数据的综合,CSA 作为数据融合的一部分——较高层次的 level 2 融合.为了实现这一目标,需要进行更广泛而全面的综合.无论是在技术级、功能级还是系统级,只有综合才能体现 CSA 的特点和优越性,也只有综合才能完成态势感知任务,实现数据融合的目标.根据以上分析我们认为,未来研究趋势重点在以下几个方面:

(1) CSA 评价体系的研究

明确统一的评价标准是开展 CSA 研究的前提,也是 CSA 研究走向成熟的标志.评价体系为 CSA 研究指明了方向,有目标、有针对性的研究才有意义.建立 CSA 评价体系至少要完成以下 3 个阶段的任务,最终实现完善、统一、可操作的评价体系:

其一,明确态势定义,达成对态势状态进行划分、定级的共识,建立形式化描述;

其二,制定评价态势评估方法准确性的度量标准,选择具体的度量指标,建立规范的准确性度量方法;

其三,以准确性为中心,综合其他系统评价指标,建立 CSA 系统评价标准.

(2) 准确而高效的评估方法的研究

制定了准确性度量标准,评估方法研究也就明确了目标,可以有的放矢,改变评估方法种类繁多、良莠不齐、存在大量重复研究的现状.从众多评估方法中选择适合 CSA 特点、准确而高效的方法,并且结合 CSA 具体问题,有针对性地进行改进和优化,在满足准确性的同时提高态势评估的效率,降低空间开销.需要指出,3 类评估方法并不是互斥的,而是从不同的问题出发寻求解决方案.如果综合使用则能取长补短.例如:证据理论与模糊逻辑相结合,能够提高处理模糊性的能力;而神经网络或者粗集理论在聚类的基础上进行将有效缩小样本空间,降低复杂性.可见,在技术级,综合数据融合和数据挖掘技术,灵活运用多种理论方法,实现从数据到信息再到知识的抽象,也是态势评估的发展方向.

(3) 网络系统表示的研究

从本质上认识网络这个复杂巨系统,是对其态势进行感知的基础.复杂系统描述本来就是难点,对于 CSA 既要借鉴现有成熟的系统表示方法,又不能机械模仿,需要具体分析网络元素及其联系,建立能够反映纷繁丰富内容和错综复杂关系的网络模型.同时,也要探索适合网络系统的创新性的表示方法,不落窠臼.

(4) 系统实际部署的研究

以 CSA 系统评价为指导,将理论研究部署到应用系统之中.在功能级,综合现有各种单元网管技术和各个领域的应用研究,研究功能模块之间的通信和交互,实现对整个网络全方位立体化动态的态势感知;在系统级,超越 CSA 本身,模糊数据融合各级之间的界限,将 level 2 融入其中,使之相互渗透,构建完整、协调的网络数据融合系统,从而在实际应用中展示 CSA 强化管理、支持决策的重要作用.

© 中国科学院软件研究所 http://www.jos.org.cn

(5) 人机交互机制的研究

人在CSA系统中始终是一个重要环节,尤其在评估阶段,很少有一种方法能够独立完成评估任务.即使像聚类这种无监督学习方法,也只能依据“异常事件很少发生,正常行为占网络行为的大部分”这一假设对态势进行简单划分.建立人机交互机制,包括以下工作:① 提供友好的人机交互界面;② 随时接收领域知识和专家建议,并及时对模型进行修改;③ 实现从自然语言到计算机能够处理的数学表达式的转换;④ 将数据挖掘技术与领域知识相结合,实现数据融合level 5的重要功能——用户细化.

7 总 结

本文在充分研究CSA国内外相关工作的基础上提出CSA问题体系,明确了CSA的研究内容;与问题体系相对应,分析了CSA的技术体系,重点介绍了模型、知识表示和评估方法3方面内容;并且讨论了CSA应用体系,包括安全、传输、生存性、系统评价等领域,每一领域涉及多方面问题,采用特殊技术解决具体问题;最后分析了未来的发展趋势.

CSA的探索才刚刚开始,与网络监测等属于level 1融合的研究相比,还处于幼年,而其困难程度却更高.这是因为网络自身灵活的组织和生长方式,增加了复杂性和不确定性,使得对网络的认知和管理变得遥不可及.与此同时,我们也应该乐观地看到,有关网络管理的研究已经比较成熟,为态势感知提供了丰富的信息;数据融合的研究也相当广泛,提出了通用的体系框架和评估模型,为态势感知奠定了理论基础.

References:

[1] Bass T. Multisensor data fusion for next generation distributed intrusion detection systems. In: Proc. of the ’99 IRIS National Symp. on Sensor and Data Fusion. Laurel, 1999. 24–27. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.1753&rep=rep1&type=ps

[2] Bass T. Intrusion systems and multisensor data fusion. Communications of the ACM, 2000,43(4):99−105. [doi: 10.1145/332051.332079]

[3] Wang J, Zhang FL, Fu C, Chen LS. Study on index system in network situation awareness. Computer Applications, 2007,27(8): 1907−1909 (in Chinese with English abstract).

[4] Ticha B, Ranchin T. A case based reasoning data fusion scheme: Application to offshore wind energy resource mapping. In: Proc. of the Int’l Conf. on Information Fusion (FUSION). 2006. 1−5. http://ieeexplore.ieee.org/

[5] Gad A, Farooq M. Data fusion architecture for maritime surveillance. In: Proc. of the Int’l Society on Information Fusion (ISIF). 2002. 448−455. http://www.isif.org/fusion/proceedings/fusion02CD/pdffiles/papers/M4D03.pd

[6] Kadar I. Knowledge representation issues in perceptual reasoning managed situation assessment. In: Proc. of the FUSION. 2005. 13−15. http://ieeexplore.ieee.org/

[7] Hall D, Llinas J. An introduction to multisensor data fusion. Proceedings of the IEEE, 1997,85(1):6−23.

[8] Blasch E, Plano S. JDL level 5 fusion model “user refinement” issues and applications in group tracking. In: Proc. of the Signal Processing, Sensor Fusion, and Target Recognition XI, SPIE Vol.4729. 2002. 270−279. http://erikblasch.tripod.com/UserRefineGroupTracking.pdf

[9] Blasch E, Plano S. DFIG level 5 issues supporting situational assessment reasoning. In: Proc. of the FUSION. 2005. 35−43. http://ieeexplore.ieee.org/

[10] Endsley M. Situayion awareness global assessment technique (SAGAT). In: Proc. of the IEEE ’88 National Aerospace and Electronics Conf. (NAECON). 1988. 789−795. http://www.satechnologies.com/services/measurement/SAGAT/

[11] Salerno J, Hinman M, Boulware D. Building a framework for situation awareness. In: Proc. of the FUSION. Stockholm, 2004. 1−8. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.63.1633&rep=rep1&type=pdf

[12] Tadda G, Salerno J, Boulwarea D, Hinmana M, Gorton S. Realizing situation awareness in a cyber environment. In: Multisensor BV, ed. Proc. of the Multisource Information Fusion, SPIE Vol.6242. 2006. 1−8. http://spiedl.aip.org/getabs/servlet/GetabsServlet?prog=normal&id=PSISDG006242000001624204000001&idtype=cvips&gifs=yes&ref=no

[13] Zhuo Y, Zhang Q, Gong ZH. Cyberspace situation representation based on niche theory. In: Proc. of the ICIA. Zhangjiajie, 2008. 1400−1405. http://ieeexplore.ieee.org/

[14] Klir G, Yuan B. Fuzzy Sets and Fuzzy Logic. New York: Prentice Hall, 1995.

[15] Chen LY, Huang J. Survey of research on measure of uncertainty. Journal of Circuits and Systems, 2004,9(3):105−111 (in Chinese with English abstract).

© 中国科学院软件研究所 http://www.jos.org.cn

[16] Grenon P, Smith B. SNAP and SPAN: Towards dynamic spatial ontology. In: Proc. of the Spatial Cognition and Computation. 2003. 137−171. http://ontology.buffalo.edu/smith/articles/SNAP_SPAN.pdf

[17] Little E, Rogova G. Ontology meta-model for building a situational picture of catastrophic events. In: Proc. of the FUSION. 2005. 796−803. http://ieeexplore.ieee.org/

[18] Chen XZ, Zheng QH, Guan XH, Lin CG. Quantitative hierarchical threat evaluation model for network security. Journal of Software, 2006,17(4):885−897 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/17/885.htm [doi: 10.1360/jos170885]

[19] Cai W. Extension Engineering Method. Beijing: Science Press, 1997 (in Chinese).

[20] Hinman M. Some computational approaches for situation assessment and impact assessment. In: Proc. of the ISIF. 2002. 687−693. http://ieeexplore.ieee.org/

[21] Jiang Y, Lin C, Wu JP. Integrated performance evaluation criteria for network traffic control. Chinese Journal of Computers, 2002, 25(8):869−877 (in Chinese with English abstract).

[22] Zhao KQ. Set Pair Analysis and Applications. Hangzhou: Science and Technology Press, 2000 (in Chinese).

[23] Das S, Lawless D, Ng B, Pfeffer A. Factored particle filtering for data fusion and situation assessment in urban environments. In: Proc. of the FUSION. 2005. 955−962. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.142.6057&rep=rep1&type=pd

[24] Zhang Y, Ji Q, Loonet C. Active information fusion for decision making under uncertainty. In: Proc. of the ISIF. 2002. 643−650. http://www.ecse.rpi.edu/homepages/qji/Papers/fusion02_zhang.pdf

[25] Mirmoeini F, Krishnamurthy V. Reconfigurable Bayesian networks for hierarchical multi-stage situation assessment in battlespace. In: Proc. of the 39th Asilomar Conf. on Signals, Systems and Computers. 2005. 104−108. http://ieeexplore.ieee.org/

[26] Shafer G. A Mathematical Theory of Evidence. Princeton: Princeton University Press, 1976.

[27] Xu XH, Liu ZL. A method for situation assessment based on D-S evidence theory. Electronics Optics &Control, 2005,12(5):36−37 (in Chinese with English abstract).

[28] Wei SZ, Zhao H, Wang G, Zhang XD. Situation assessment model of complex system and its implementation method based on ontology. Journal of System Simulation, 2005,17(5):1200−1202 (in Chinese with English abstract).

[29] Li WS, Wang BS. A synthetic method for situation assessment based on fuzzy logic and D-S evidential theory. Systems Engineering and Electronics, 2003,25(10):1278−1280 (in Chinese with English abstract).

[30] Deng JL. Gray Control System. Wuhan: Publishing House of Center-China University of Technology, 1985 (in Chinese).

[31] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Boston: Kluwer Academic Publishers, 1991. 1−10.

[32] Wei SZ, Jin ND, Hui XJ, Liu H, Zhang XD. A situation assessment model and its application based on data mining. In: Proc. of the FUSION. 2006. 1−7. http://ieee-aess.org/isif/sites/default/files/proceedings/fusion06CD/Papers/322.pdf

[33] Zhuo Y, Zhang Q, Gong ZH. Network situation assessment based on RST. In: Proc. of the PACIIC. Wuhan, 2008. 502−506. http://ieeexplore.ieee.org/

[34] Lakhina A, Crovella M, Diot C. Mining anomalies using traffic feature distributions. In: Proc. of the ACM SIGCOMM. 2005. 217−228. http://www.sigcomm.org/sigcomm2005/paper-LakCro.pdf

[35] Wang H, Gong ZH. Algorithm based on entropy for finding critical traffic matrices. Journal of Software, 2009,20(5):1377−1383 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3308.htm [doi: 10.3724/SP.J.1001.2009.03308]

[36] Kosala R, Blocked H. Web mining research: A survey. ACM SIGKDD Explorations, 2000,2(1):1−15. [doi: 10.1145/360402.360406]

[37] Dhillon I, Guan Y, Kogan J. Iterative clustering of high dimensional text data augmented by local search. In: Proc. of the 2002 IEEE Int'l Conf. on Data Mining. 2002. 131−138. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.8317&rep=rep1&type=pdf

[38] Zhuo Y, Zhang Q, Gong ZH. Research and implementation of network transmission situation awareness. In: Proc. of the CSIE. Los Angeles, 2009. 210−214. http://ieeexplore.ieee.org/

[39] Sun JG, Liu J, Zhao LY. Clustering algorithms research. Journal of Software, 2008,19(1):48−61 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/19/48.htm [doi: 10.3724/ SP.J.1001.2008.00048]

[40] Bass T, Robichaux R. Defense-in-Depth revisited: Qualitative risk analysis methodology for complex network-centric operations. In: Proc. of the Communications for Network-Centric Operations: Creating the Information Force (MILCOM). IEEE, 2001. 64−70. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.65.5445&rep=rep1&type=pdf

[41] Wei Y, Lian YF, Feng GD. A network security situational awareness model based on information fusion. Journal of Computer Research and Development, 2009,46(3):353−362 (in Chinese with English abstract).

[42] Wei Y, Lian YF. A network security situational awareness model based on log audit and performance correction. Chinese Journal of Computers, 2009,32(4):763−772 (in Chinese with English abstract).

© 中国科学院软件研究所 http://www.jos.org.cn

[43] Liu DP, Fei AG, Li G. Research on C4ISR information superiority measurement. Computer Simulation, 2007,24(6):27−30 (in Chinese with English abstract).
[44] Zhu L, Wang HQ, Zheng LJ. Survey of network security situation visualizations. 2006. http://www.paper.edu.cn
[45] Lau S. The spinning cube of potential doom. 2003. http://www.nersc.gov/nusers/security/TheSpinningCube.php
[46] Yang YH, Li XD. The study of a framework for IP network performance metrics. Journal on Communications, 2002,23(11):1−7 (in Chinese with English abstract).
[47] Lin C, Zhou WJ, Tian LQ. Research on performance evaluation criteria for IP network traffic control. ACTA Electronica Sinica, 2002,30(12A):1973−1977 (in Chinese with English abstract).
[48] Zhang DY, Hu MZ, Zhang HL. Study on network performance evaluation method based on measurement. Journal on Communications, 2006,27(10):74−79 (in Chinese with English abstract).
[49] Jiang XP. Design and realization of an integrated evaluation method of network performance. Journal of Naval University of Engineering, 2006,18(5):74−78 (in Chinese with English abstract).
[50] Blasch E, Pribilski M, Daughtery B, Roscoe B, Gunsett J. Fusion metrics for dynamic situation analysis. In: Kadar I, ed. Proc. of the Signal Processing, Sensor Fusion, and Target Recognition XIII, SPIE Vol.5429. Bellingham, 2004. 428−438. http://spiedl.aip.org/getabs/servlet/GetabsServlet?prog=normal&id=PSISDG005429000001000428000001&idtype=cvips&gifs=yes&ref=no
[51] Salernoa J, Blasch E, Hinmana M, Boulwarea D. Evaluating algorithmic techniques in supporting situation awareness. In: Multisensor BV, ed. Proc. of the Multisource Information Fusion: Architectures, Algorithms, and Applications 2005, SPIE Vol.5813. Bellingham, 2005. 96−104. http://adsabs.harvard.edu/abs/2005SPIE.5813...96S

附中文参考文献:

[3] 王娟,张凤荔,傅翀,陈丽莎.网络态势感知中的指标体系研究.计算机应用,2007,27(8):1907−1909.
[15] 陈理渊,黄进.不确定度问题研究情况综述.电路与系统学报,2004,9(3):105−111.
[18] 陈秀真,郑庆华,管晓宏,林晨光.层次化网络安全威胁态势量化评估方法.软件学报,2006,17(4):885−897. http://www.jos.org.cn/1000-9825/17/885.htm [doi: 10.1360/jos170885]
[19] 蔡文.可拓工程方法.北京:科学出版社,1997.
[21] 江勇,林闯,吴建平.网络传输控制的综合性能评价标准.计算机学报,2002,25(8):869−877.
[22] 赵克勤.集对分析及其初步应用.杭州:浙江科技出版社,2000.
[27] 徐晓辉,刘作良.基于 D-S 证据理论的态势评估方法.电光与控制,2005,12(5):36−37.
[28] 魏守智,赵海,王刚,张晓丹.复杂系统态势评估模型及其本体论实现方法.系统仿真学报,2005,17(5):1200−1202.
[29] 李伟生,王宝树.基于模糊逻辑和 D-S 证据理论的一种态势估计方法.系统工程与电子技术,2003,25(10):1278−1280.
[30] 邓聚龙.灰色控制系统.武汉:华中理工大学出版社,1985.
[35] 王宏,龚正虎.一种基于信息熵的关键流量矩阵发现算法.软件学报,2009,20(5):1377−1383. http://www.jos.org.cn/1000-9825/3308.htm [doi: 10.3724/SP.J.1001.2009.03308]
[39] 孙吉贵,刘杰,赵连宇.聚类算法研究.软件学报,2008,19(1):48−61. http://www.jos.org.cn/1000-9825/19/48.htm [doi: 10.3724/SP.J.1001.2008.00048]
[41] 韦勇,连一峰,冯国登.基于信息融合的网络安全态势评估模型.计算机研究与发展,2009,46(3):353−362.
[42] 韦勇,连一峰.基于日志审计与性能修正算法的网络安全态势评估模型.计算机学报,2009,32(4):763−772.
[43] 刘东坡,费爱国,李革.C4ISR 系统信息优势度量研究.计算机仿真,2007,24(6):27−30.
[44] 朱亮,王慧强,郑丽君.网络安全态势可视化研究评述.中国科技论文在线,2006. http://www.paper.edu.cn
[46] 杨雅辉,李小东.IP 网络性能指标体系的研究.通信学报,2002,23(11):1−7.
[47] 林闯,周文江,田立勤.IP 网络传输控制的性能评价标准研究.电子学报,2002,30(12A):1973−1977.
[48] 张冬艳,胡铭曾,张宏莉.基于测量的网络性能评价方法研究.通信学报,2006,27(10):74−79.
[49] 蒋序平.网络性能综合评估方法 IEMoNP 的设计和实现.海军工程大学学报,2006,18(5):74−78.

龚正虎(1945－),男,湖南长沙人,教授,博士生导师,主要研究领域为计算机网络.

卓莹(1979－),女,博士生,主要研究领域为态势感知,网络管理.

© 中国科学院软件研究所 http://www.jos.org.cn