

面向行为的内网空间安全态势建模与分析

朱承¹, 缪嘉嘉², 毛捍东³

(1. 国防科学技术大学信息系统工程重点实验室, 湖南 长沙 410073;
2. 解放军理工大学指挥信息系统学院, 江苏 南京 210007;
3. 北京普世时代科技有限公司, 北京 100083)

摘要: 依据态势的一般理论, 结合内网安全的特点, 将内网空间划分为物理域、信息域、社会域, 提出了面向行为的内网安全态势——PMDV 模型。该模型抽象出内网空间物理域和信息域中人(P)、机(M)、数据(D)、虚拟实体(V)等4类行为主体, 各类主体及其交互定义态势中的核心要素——行为, 为面向行为的内网空间安全态势构建提供了依据。在此基础上, 进一步给出了面向行为的内网空间安全态势分析平台的结构设计及其实现途径。

关键词: 网络空间; 内网; 安全; 态势; 行为

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2016.00067

Behavior-oriented inside network security situation modeling and analysis

ZHU Cheng¹, MIAO Jia-jia², MAO Han-dong³

(1. Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China;
2. School of C4ISR, PLA University of Science and Technology, Nanjing 210007, China;
3. Pushtime Science and Technology Inc., Beijing 100083, China)

Abstract: According to the situation awareness theory as well as the characteristics of the inside network security, a PMDV model for behavior-oriented inside network security situation modeling by dividing the inside network space into physical domain, information domain and social domain was proposed. A sound base for situation modeling of inside network security, which classified the entities inside the network space into the following 4 classes: person(P), machine(M), data(D) and virtual agent(V), was provided and the core element “behavior” in situation model with the interactions between different PMDV entities was defined. The structure and implementation guideline of the situation analysis platform based on PMDV model were also presented and studied.

Key words: cyberspace, inside network, security, situational awareness, behavior

1 引言

内网安全是网络安全中的重要领域, 对保密单位而言尤为重要。但是, 随着智能移动设备、物联网和云计算的快速发展和不断应用, 传统的以合规性检查、漏洞封堵为主的静态内网安全防

护理念和技术在新条件下暴露出大量问题。例如, 以特征匹配为主的安全手段、设备对 0-day 等未知漏洞或威胁无能为力; 电磁、无线、电力等渗透手段层出不穷, 物理隔离并不能保证网络的封闭性; 最大的危险仍然来自于组织内部人员无意或有意的泄密行为。IBM 发布的 2015 年赛博安

收稿日期: 2016-06-08; 修回日期: 2016-07-02。通信作者: 朱承, zhucheng@nudt.edu.cn

基金项目: 总装预研基金资助项目 (No.9140A15070414JB25224)

Foundation Item: The General Equipment Department Foundation (No.9140A15070414JB25224)

全情报表明,约55%的攻击来自内部^[1]。

与此同时,信息技术的飞速发展和越来越广泛的应用使组织内部的网络和信息系统每天都在快速产生大量的机器数据。大型企业的信息系统每天产生的日志事件数量可以达到100亿到1000亿条^[2],中小规模的组织或企业内部产生的机器数据也达到可观的规模。这些数据记录了组织内部用户各类活动和交互的完整行为,具有海量、高速、多样、低价值密度等明显的大数据特点,有巨大的分析和利用价值。如何有效利用这些数据成为新时期需要面对和解决的问题。

面对网络安全中的新形势、新问题,除了更严格的加密和更细粒度的认证与授权外^[3],数据驱动的动态综合安全理念已成为网络安全解决思路的趋势和潮流^[2,4],基于大数据、行为分析的新型智能化安全管理工具大量涌现^[5-7]。内网安全管理需要依靠对海量数据的深度挖掘与机器学习,高效、准确地发现异常、危险行为,才能有效地应对高级可持续性威胁(APT, advanced persistent threat)和来自内部人员的威胁,实现由“被动”到“主动”的信息安全防护。因此,如何针对内网特点和行为分析的需求,构建内网空间安全态势模型,是个重要的问题。

2 面向行为的内网空间安全态势模型

2.1 内网空间

为了刻画人类生存的信息环境或信息空间,人们创造了Cyberspace一词。2008年,美国“54号国家安全总统令”对Cyberspace的定义为^[8]:Cyberspace是信息环境中的一个整体域,由独立且互相依存的信息基础设施和网络组成,包括互联网、电信网、计算机系统、嵌入式处理器和控制器系统。Cyberspace对应的中文名称有多种,如网络空间、网电空间、赛博空间等,本文采用网络空间这一名称。网络空间已成为由计算机、网络构成的信息社会的代名词。

本文的内网空间是指一个组织内部的网络空间。与网络空间相比,内网空间的特点主要体现在以下3个方面。

1) 内网有对应的组织结构以及相对明确的物理边界。

2) 内网,特别是有保密需要的内网,往往存在多套不同密级网络。多个网络之间物理隔离或者通过安全交换设备(单向)连接。

3) 物理空间是内网安全中的重要方面。

参考网络空间的定义并结合内网特点,本文将内网空间划分为以下3个域。

1) 物理域

物理域由内网对应的物理边界中的人、机器设备等实体构成,是内网中物理实体所存在的物理空间。

2) 信息域

信息域是信息获取、传输、加工、处理、存储、显示和共享的空间,是内网中信息驻留、流动、被加工、被使用的“生态环境”。

3) 社会域

社会域包括内网所对应组织中的组织结构、交互关系、业务关系等。

内网空间域层次模型如图1所示。

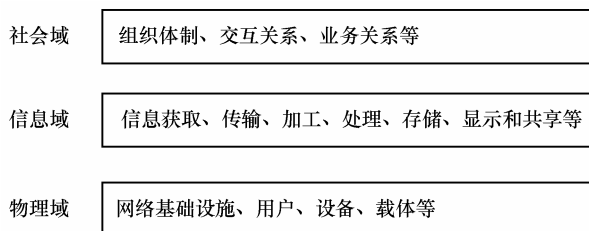


图1 内网空间域层次模型

2.2 PMDV 态势模型

网络空间安全^[8]包括以下几个方面。

1) 设备安全: 信息系统的设备安全是信息系统安全的物质基础。

2) 数据安全: 确保数据免受未授权的泄露、篡改和毁坏。

3) 内容安全: 指信息内容在政治、法律、道德层次上的要求,如政治健康、符合法律等。

4) 行为安全: 行为的过程和结果不能危害数据的完整性,当行为的过程出现偏离预期时,能够发现、控制或纠正。

由内网的特点可知,内网空间安全的核心是数据安全,数据安全实现的支撑是行为安全。

“态势”一词来源于作战指挥,是指指挥控制的核心要素,与决策密切相关。态势指状态与形势,包括人或事表现出的形态、环境或事物的发展状

况和情形等^[9]。

安全态势建模对于安全至关重要。已有的网络空间安全态势建模的研究,借鉴了态势感知的一般理论,对于态势本身的建模,则主要从静态合规性检查、漏洞评估等角度研究态势评估的指标体系。例如,文献[10,11]从信息资产、脆弱性指数、威胁性指数、风险程度等方面归纳了安全态势指标体系。但是,这一类基于指标的安全态势建模方法缺乏动态角度的行为刻画,对于理解网络空间中的实体及其交互关系缺乏指导意义。在内部威胁的研究中,不少文献从概念、过程建模等角度提出了相应的模型。例如,文献[12]提出内部威胁的形式化定义和分类,文献[13]提出内部人在事件触发下的恶意行为过程模型。这些模型主要面向概念理解和过程解释,不涉及态势要素本身的建模与抽象。有些研究给出了与行为相关的态势要素,如文献[14]从文件访问记录、USB设备使用记录等方面给出内部人与安全相关的行为要素,但缺乏建模的整体依据和框架。针对以上问题,本文从动态行为分析的角度,提出内网安全态势建模框架,为安全态势的抽象提供依据。

在态势模型中,态势元素被抽象为环境、实体、事件、组及行为5个方面^[15]。在一定环境中存在的某事物(实体),当其性质或状态发生改变时,就产生了事件。每个实体所发生的事件在时间和空间上都存在某种联系,多个实体和事件按照某种关系组合在一起构成组。在特定环境中,组或单个实体会产生某种行为,态势则被建模为这一过程的描述。因此,态势2个最活跃的元素是实体和行为。

要建立面向行为分析的态势模型,首先要明确行为的主体。基于第2.1节中对内网空间的域

划分,本文通过抽象物理域、信息域中的实体类型,将内网行为主体分为以下4类,简称PMDV。

1) 人(P)

人是内网中最关键的行为主体。这类实体包括个人、特征人群、组织等。

2) 机(M)

机指内网中的各类机器设备,包括主机、服务器、交换机、网络设备、移动载体等硬件设备。

3) 数据(D)

数据指内网中的各类数据,是防泄密的核心对象。如用户数据(文本、图片、视频)、系统数据、网络数据等。

4) 虚拟实体(V)

虚拟实体指网络空间中的各类虚拟实体,如各类系统进程、服务进程、恶意进程(病毒、木马)等。

在PMDV中,人、机属于物理域的实体;数据、虚拟实体则属于信息域的实体。PMDV行为主体如图2所示。

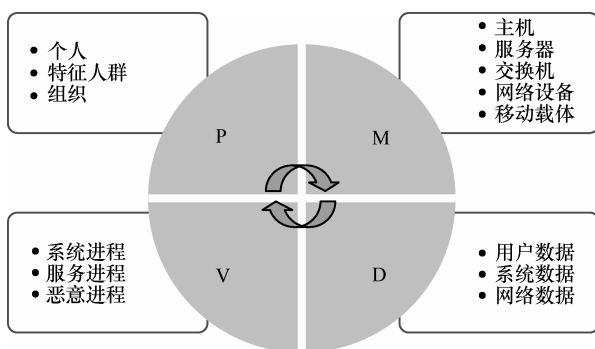


图2 内网空间行为主体

根据态势模型中对环境、实体、事件、组及行为等态势元素的抽象,本文构建面向行为的PMDV内网空间安全态势模型,如图3所示。其中,行为主体以及对应的行为由PMDV4类主体

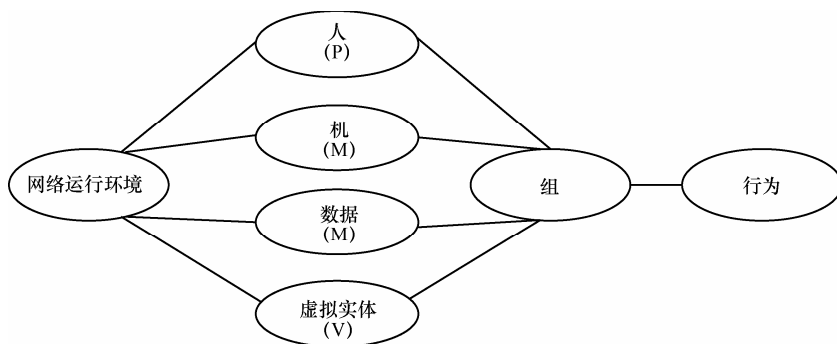


图3 内网空间安全态势元素及关系

及其相互之间的交互关系定义：PMDV 中某一类型的实体或者多类实体形成的组在网络运行环境中的活动构成了内网空间中的行为，网络运行环境指网络拓扑、参数等网络运行状态。

PMDV 模型提供了内网空间安全态势中行为建模的依据。图 4 给出了 PMDV 模型中实体及两两交互所定义的行为示例（由于不区分交互关系中的序，图 4 中 PMDV 交互矩阵为对称阵，左下角矩阵留空），其中，“/”表示无交互实体。

3 面向行为的内网安全态势生成与分析平台

面向行为的态势生成与分析必须基于行为相关数据，并依托相应的技术平台。数据的多源异构性、规模以及分析需求决定了采用大数据技术构建分析平台的合理性，即以大数据技术为支撑，构建态势生成和分析平台，实现行为数据采集、存储、关联与融合；安全态势的感知和可视化；基于机器

	人	机	数据	虚拟实体	/
人	谁与谁经常交互？交互的规律是什么？	谁使用哪台机器/载体/设备？在哪里使用？有什么规律？	谁访问什么数据？在哪台机器/载体上访问？访问的规律是什么？	谁在运行什么程序？谁带进了病毒/木马？	何时进/出大楼和房间？在哪些区域活动？
机	/	客户机和服务器之间的交互有什么规律？移动载体经常在哪些主机上使用？	主机/服务器/移动载体上存取了什么数据？规律是什么？	主机/服务器运行了哪些程序？恶意进程在主机/服务器间的传播与扩散规律是什么？	网络中有哪些主机/服务器/移动载体？在什么地方？
数据	/	/	数据之间的复制、转换、传输的关系和规律是什么？	各类进程产生了哪些数据？哪些数据被（恶意）进程感染/影响/访问？	数据本身有什么特点和规律？
虚拟实体	/	/	/	进程之间的交互行为规律是什么？重要的服务进程被哪些客户进程访问？	网络中存在哪些活跃的进程？活动规律是什么？存在哪些恶意进程？

图 4 PMDV 模型中各类实体及两两交互所定义的行为示例

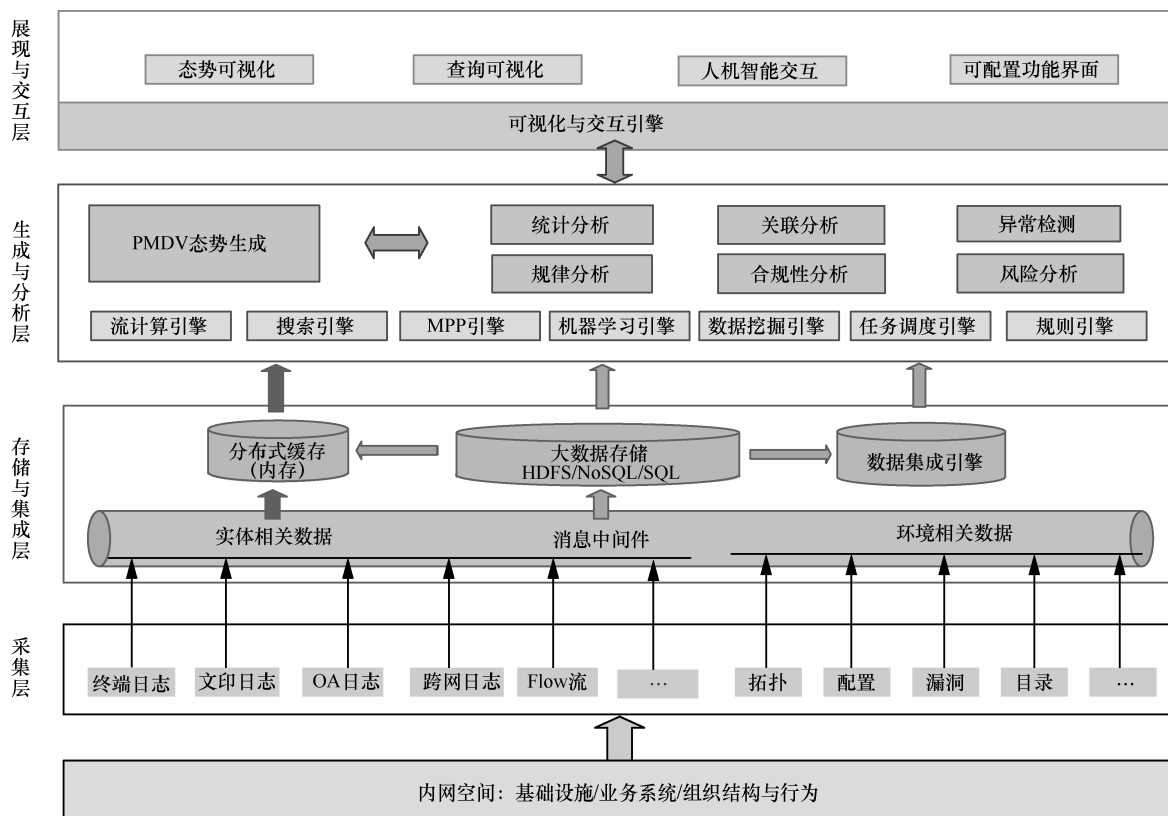


图 5 面向行为的内网空间安全态势分析平台架构

学习、数据挖掘等技术实现智能化安全态势分析能力; 实现行为画像和异常检测, 为发现未知危险、APT 攻击、内部人员的恶意行为等提供基础。

态势生成与分析平台分为采集层、存储与集成层、生成与分析层、展现与交互层, 架构如图 5 所示。

1) 采集层

采集层从内网空间采集与安全态势相关的大量异构数据, 主要分为 2 大类型数据。一类为内网中各类实体及其行为相关数据, 以海量、高速、异构为特征, 数据源主要包括终端/服务器日志、业务系统日志、网络流日志、跨网交换日志、门禁、监控视频等; 另一类为网络基础环境数据, 包括网络拓扑信息、配置信息、漏洞信息、用户目录等。在线数据采集可使用 LogStash^[16]等数据采集系统, 实现对数据的采集和预处理。对于隔离网络中的数据源, 可采用离线方式进行批量导入。

2) 存储与集成层

存储与集成层基于大数据技术实现对采集数据的预处理和存储。采集的各类数据, 先进入消息中间件(如 Kafka^[17]), 以解决数据的缓存和容错问题。数据存储系统从消息中间件获取采集数据进行存储, 并综合利用各类 NoSQL 技术^[18]提供的高扩展性实现海量数据的存储和管理。例如, 日志等半结构化信息可以使用 Elastic Search^[16]等分布式文档数据库进行存储, 实现字段解析和全文检索; 图像、视频等非结构化数据可基于 HDFS^[19]分布式文件系统进行存储和管理。数据集成可采用 Hive^[20]等数据引擎, 实现对多源数据的统一访问和查询。需要实时处理的数据可以直接从消息中间件进入基于内存的缓存系统, 供在线分析和计算。

3) 生成与分析层

生成与分析层基于 PMDV 模型实现对内网空间安全态势的计算与生成, 并支持对态势及态势相关数据进行实时和历史分析。对态势的分析包括统计分析、行为规律分析、关联分析、合规性分析、异常检测、风险分析等。态势生成与分析可以基于 Hadoop、Spark 等开源大数据社区提供的多种大数据分析模式和大量常用的机器学习模型实现。例如, Spark^[21]以容错分布式内存数据结构 RDD^[22]为基础, 同时支持

MapReduce 方式的数据批处理分析、迭代式图分析和在线流分析等多种分析模式, 并提供 MLlib 机器学习算法库。

4) 展现与交互层

展现与交互层对态势进行可视化展现, 并实现用户与态势的交互。用户可以从 PMDV 各维度以及维度的组合观察、感知内网空间行为, 并可进行 ad hoc 方式的数据查询与交互。主流的大数据可视化工具, 如 Kibana^[16]等, 均支持数据分析指标及展现方式的定制与交互, 可以作为实现的基础。

5 结束语

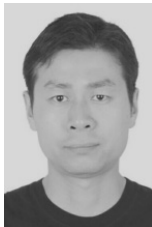
PMDV 模型抽象出了内网空间物理域和信息域中人、机、数据、虚拟实体等 4 类行为主体, 通过各类主体及其交互定义了态势核心要素——行为, 为面向行为的内网空间安全态势构建提供了依据。PMDV 态势模型的后续研究包括进一步考虑内网空间的电磁环境, 以及态势分析平台的实现与实验等。

参考文献:

- [1] IBM security. IBM 2015 cyber security intelligence index[R]. IBM, 2015.
- [2] CARDENAS A A, MANADHATA P K, RAJAN S. Big data analytics for security[J]. IEEE Security & Privacy, 2013, (11/12): 74-76.
- [3] RORY W, BETSY B. BeyondCorp: a new approach to enterprise security[J]. Login, 2014, 39(6):6-11.
- [4] JAY J. Data driven security: analysis, visualization and dashboards[M]. Indianapolis: John Wiley and Sons, 2014.
- [5] Splunk Inc. Using splunk user behavior analytics[R]. Splunk, 2015.
- [6] 启明星辰. 基于大数据分析的安全管理平台技术研究及应用[EB/OL]. <http://www.venustech.com.cn/NewsInfo/531/25566.Html>. 2014.
- [7] Venustech. Study and application of information security management platform based on big data[EB/OL]. <http://www.venustech.com.cn/NewsInfo/531/25566.Html>. 2014.
- [7] 瀚思. 数据驱动安全[EB/OL]. http://www.hansight.com/HanSight_Intro_2015V2.pdf. 2015.
- [8] 张焕国, 韩文报, 来学嘉, 等. 网络空间安全综述[J]. 中国科学, 2016, 46(2):125-164.
- [9] ENDSLEY M R. Toward a theory of situation awareness in dynamic systems[J]. Human Factors, 1995, 37(1): 32-64.
- [10] 尤马彦. 网络安全态势评估技术的研究与实现[D]. 广东工业大学, 2012.

- YOU M Y. Research and implementation of network security situation assessment technology[D]. Guangzhou: Guangdong University of Technology, 2012.
- [11] 王志平. 基于指标的网络安全态势评估研究[D]. 国防科学技术大学, 2010.
- WANG Z P. Network security situation assessment research based on the index[D]. Changsha: National University of Defense Technology, 2010.
- [12] 杨光, 马建刚, 于爱民, 等. 内部威胁检测研究[J]. 信息安全学报, 2016,1(3): 21-36.
- YNAG G, MA J G, YU A M, et al. The internal threat detection study[J]. Journal of Information Security, 2016,1(3): 21-36.
- [13] NURSE J R C, BUCKLEY O, LEGG P A, et al. Understanding insider threat: a framework for characterising attacks[C]//IEEE Symposium on Workshop on Research for Insider Threat, IEEE Computer Society Security & Privacy. c2014 :215-228.
- [14] AZARIA A, RICHARDSON A, KRAUS S, et al. Behavioral analysis of insider threat: a survey and boot- strapped prediction in imbalanced data[J]. IEEE Transactions on Computational Social Systems, 2014, 1(2):135-155.
- [15] JAKOBSON G, BUFORD J, LEWIS L. A framework of cognitive situation modeling and recognition[C]// 2006 IEEE Military Communications Conference. c2006:1-7.
- [16] Elastic search site[EB/OL]. <https://www.elastic.co/>.
- [17] Kafka site[EB/OL]. <http://kafka.apache.org/>.
- [18] CATTELL R. Scalable SQL and NoSQL data stores[J]. Sigmod Rec, 2011, 39(4): 12-27.
- [19] HDF site[EB/OL].https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html.
- [20] HIVE site[EB/OL]. <https://hive.apache.org/>.
- [21] Spark site[EB/OL]. <http://spark.apache.org/>.
- [22] ZAHARIA M, CHOWDHURY M, DAS T, et al. Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing[C]//The 9th USENIX Conference on Networked Systems Design and Implementation. c2012.

作者简介



朱承 (1976-), 男, 湖南长沙人, 博士, 国防科学技术大学研究员、硕士生导师, 主要研究方向为指挥控制、智能决策、分布式系统。



缪嘉嘉 (1980-), 男, 江苏如皋人, 博士, 解放军理工大学讲师, 主要研究方向为数据安全、安全大数据、人工智能。



毛捍东 (1979-), 男, 湖南岳阳人, 博士, 北京普世时代科技有限公司高级工程师, 主要研究方向为信息安全、云计算和系统仿真与决策技术。