

基于情景感知的企业级安全大数据平台建设

采用大数据技术构建基于情景感知的企业级安全大数据平台，更精准的发现高级持续威胁，从而确保安全预警的前摄性和准确性，推动企业主动防御体系建设。

杨维永¹ 王红凯² 靳丹³ 刘昀⁴/1. 南京南瑞集团公司 2. 国网浙江省电力公司信息通信分公司 3. 国网甘肃省电力公司信息通信公司 4. 国网冀北电力有限公司



杨维永/工程师

关键词/Keywords

- 安全大数据 ·
- 数据挖掘 ·
- 主动防御 ·
- 威胁情报 ·
- 安全运营中心 ·

0 引言

信息安全体系结构的设计并没有严格统一的标准，不同领域不同时期，人们对信息安全的认识不尽相同，对解决信息安全问题的侧重也各有差异。早期人们对信息安全体系的关注焦点，即以防护技术为主的静态的信息安全体系。随着人们对信息安全认识的深入，其动态性和过程性的发展要求愈显重要。P2DR安全模型^[1]和IATF信息保障技术框架^[2]是信息安全体系发展的重要里程碑，这两个模型都是在整体安全策略的控制和指导下，将人、操作和技术融合组成一个“完整的、动态的”安全威胁响应体系。Cisco的自防御网络、Enterasys的动态入侵响应系统及华为3COM 安全渗透网络等都相继提出了主动防御的概念，然而快速、准确地发现威胁是建立主动防御的基础，没有好的检测技术做保证，主动防御是不可能实现的。如何在恶意攻击行为发生前就主动检测网络脆弱环节，研究并预测攻击者行为，建立主动型防护体系是安全领域中的一个新课题。

随着黑客攻击技术不断提高，特别是APT（高级持续威胁）的出现，网络

安全问题日益突出。相对于其他攻击形式，APT攻击原理更为高级和先进，主要体现在APT攻击者在发动攻击之前需要针对攻击对象的业务流程和目标系统进行精确和详细的信息收集，收集过程中主动挖掘被攻击对象系统和应用程序的漏洞，并利用这些漏洞组建攻击者所需的网络，利用0day漏洞实施攻击。APT攻击具备攻击更快速、手段更专业、过程更隐蔽且技术更复杂的特点，这些对网络安全监测及主动防御提出了更高要求。因此，需要利用大数据分析及安全取证技术，灵活地对日志、告警和流量数据进行威胁建模和关联分析，快速计算，从海量日志告警信息中发现有效信息，定位安全威胁和脆弱性，设计特定的安全监测场景是有效开展安全监测和应急响应工作的关键；同时需建设支持灵活数据分析和精确取证的支撑性工具，提升对监测数据的深度分析能力和对安全问题的快速处置能力。

1 基于情景感知的新一代主动防御体系

1.1 APDR主动防御模型

相对被动防御（Passive Defense），主动防御（Active Defense）是一种前摄性防御，由于一些防御措施的实施，使攻击者无法完成对目标的攻击，或者使系统能够在无需人为被动响应的情况下预防安全事件。主动防御可以预测未来的攻击形势，检测未知的攻击，从根本上改变了以往防御落后于攻击的不利局面；具有自学习的功能，可以实现对网络安全防御系统进行动态的加固^[4]；主

动防御系统能够对网络进行监控，对检测到的网络攻击进行实时响应。这种响应包括牵制和转移黑客的攻击，对黑客入侵方法进行技术分析，对网络入侵进行取证，对入侵者进行跟踪甚至反击等。

情景感知 (Context Awareness) 技术非常适用于信息安全的主动防御体系，其源于所谓普适计算的研究，最早于1994年提出。情景感知简单说就是通过传感器及其相关的技术使计算机设备能够“感知”到当前的情景。情景感知计算的应用可以通过传感器获得关于用户所处环境的相关信息，从而进一步了解用户的行为动机等。在信息安全领域基于情景感知就是在特定功能的网络应用中能够识别主体、客体以及主动对客体的动机。在实际工作中情景感知基于当前条件的信息（如地点、时间、漏洞状态等），使用额外信息提升信息安全决策正确性。对于分辨哪些攻击逃过传统安全防护机制，以及帮助确定有意义的偏离正常行为而不需要增加大量误报率时，对场景的利用非常关键^[5]。

本文基于P2DR安全模型进行延伸，提出基于情景感知的主动防御模型APDR，即情景感知 (Aware) – 动态防护 (Protect) – 深度监测 (Detect) – 研判处置 (Response)，模型如图1所示。

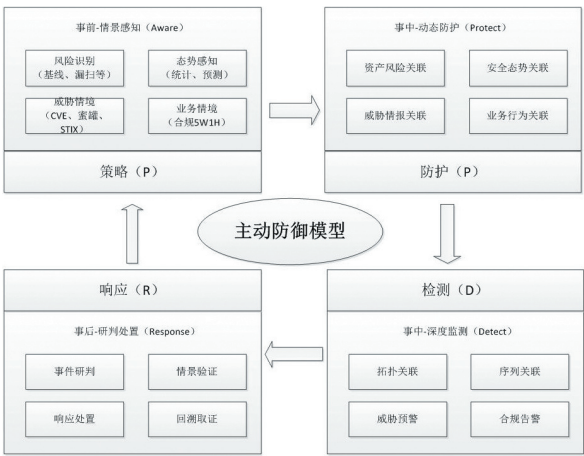


图1 基于情景感知的主动防御模型APDR

1.2 企业级主动防御体系

建立主动防御系统是一项十分复杂的工程，从IATF安全保障体系来看，涉及人 (Person)、流程 (Procedure) 和技术 (Technology) 等要素的有机结合，构成一个动态的多层次的防御系统，系统各部分应能相互协调、协作形成一个整体，完成威胁主动防御的自动管理。

在主动防御体系建设方面，为体现主动防御体系的前摄性、自学习和实时响应，除了组织流程建设外，需要在威胁感知基础设施方面不断完善，并不断完善防护、监测等技术措施，提升主动防御体系的完整性。基于情景感知的主动防御模型，本文提出企业级主动体系，如图2所示。

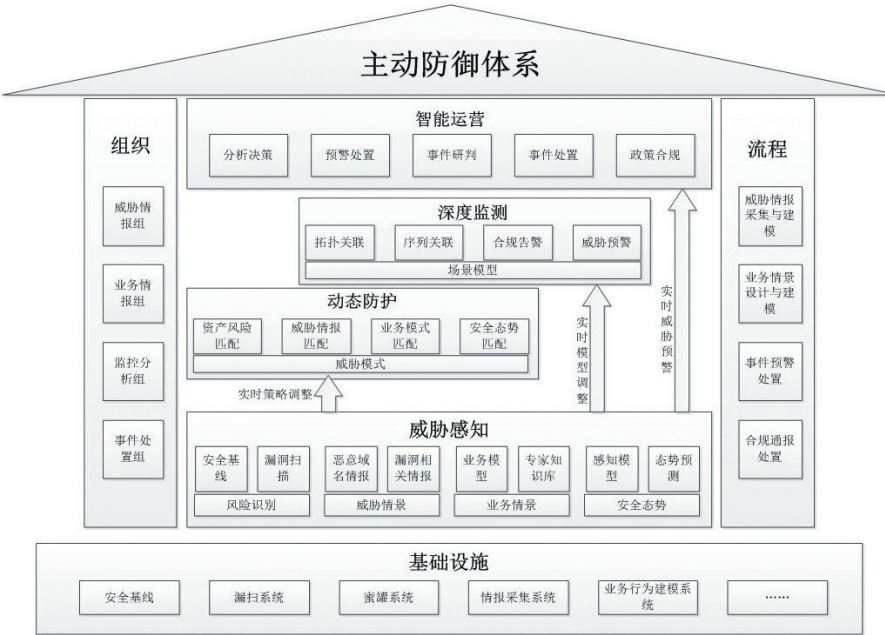


图2 基于情景感知的主动防御体系

在组织方面，需要构建内外部威胁情景信息搜集队伍，以及可以编写内外部威胁模型、对安全事件精准研判的专业人员。另外对于情报的搜集要建立适当的渠道，如国家相关部门，确保情报准确性、权威性。目前国内各大企业侧重体系建设，技术和产品的实施，对运维能力和组织建设不够，导致很多宣称已经实现主动防御体系仅落实在字面上。

与组织对应的流程体系建设，主动防御体系基于P2DR动态防御体系之上，需要组织不断完善流程，应对不断演进的内外外部威胁。建立高效的流程体系，可以确保各种技术、管理手段得以落实，从而更好地实现企业政策合规、生产经营需要。

在技术层面，构建主动防御体系需要企业规划有一个总体的、清晰线路，各种技术手段相互关联，相互补充。主动防御前摄性的前提是对网络及应用环境资产属性及安全属性有个全面的掌握，能实时发现系统的弱点、威胁和风险。在基础设施方面，安全基线系统建设最为关键，由于被防护系统会不停地进行变动，新的系统也将不断被引入，用户账户不停的新建和撤销，新的漏洞不断地披露，新应用部署；针对新威胁的适应改造也一直进行着。所以，我们也应该持续对终端设备、服务器端系统、漏洞和网络交互接口、业务交互行为进行重定基线以及挖掘发现。主动防御体系关键数据模型，如图3所示。

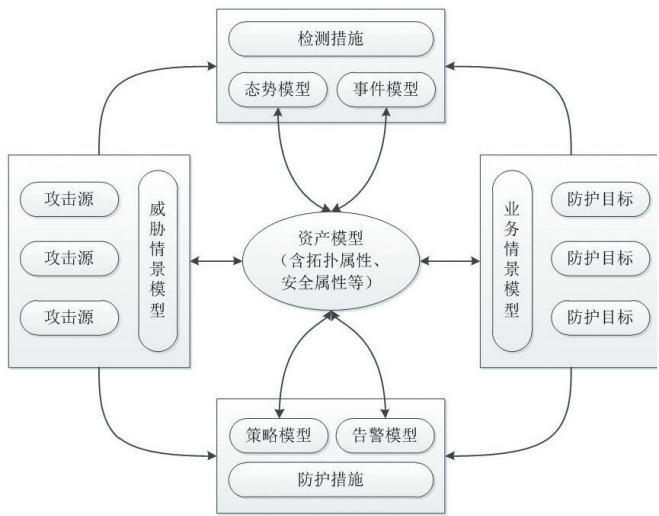


图3 主动防御体系关键数据模型

在数据模型方面，需要对主动防御体系的内涵进行诠释，确保该体系建设过程中数据流和信息流能够统一和集成。主动防御体系涉及的关键信息模型，包括威胁情报模型（需要对不同攻击源进行统一描述）、防护设备策略模型（不同厂商策略统一便于威胁情报模型在防护设备中得以动态实施）、防护设备告警模型（需要统一或规范各厂商的设备告警内容一致性）、业务情景模型（需要从系统层面、网络层面、业务层面构建统一的业务模型），以及监测层面的事件模型（统一安全事件，便于事件的统计分析和风险计算）、态势感知预测模型（对网络流量、业务交互频率、安全事件发生频率等进行建模）等。但这些模型核心是防护对象的资产模型，包括拓扑属性、安全属性和运行属性等，是上述模型的纽带和计算基础。

2 基于情景感知的安全大数据平台

2.1 安全大数据平台

传统企业在构建安全防护体系过程中，SIEM或SOC是安全监控的主要工作，但面对海量安全数据或者业务数据，性能问题凸显，且分析计算模型比较单调，无法高效、实时挖掘出安全的有效事件。思科2014年发布开源安全大数据平台OpenSOC，给SIEM或SOC的发展提供了参考性的设计。OpenSOC是一个针对网络包和流的大数据分析框架，它是大数据分析与安全分析技术的结合，能够实时检测网络异常情况并且扩展多个节点，它的存储采用开源项目Hadoop，实时索引采用开源项目Elastic Search，在线流分析采用著名的开源项目Storm。但其框架在企业实际落地有很多问题，没有融合企业安全管控流程，缺乏安全规划、安全建设、安全监测及合规判断等内容，更多是一种单纯的安全分析平台。

本文提出的企业级信息安全主动防御体系中，结合了具体的安全监测和管控应用，建立基于OpenSOC演进的企业级安全大数据平台，全面支撑企业级安全主动防御体系建设，平台结构如图4所示。平台作为贯穿安全措施（图4下方的安全数据源）和安全管理（图4上方的安全应用），采用大数据技术构建安全数据中心、计算中心和

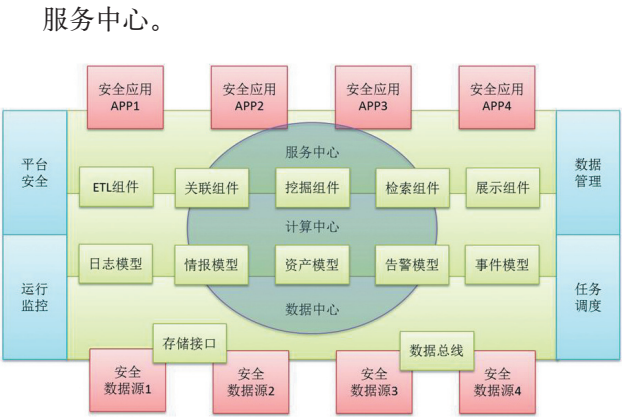


图4 企业级安全大数据平台结构

- 1) 安全数据中心。针对前文描述的主动防御体系关键数据进行保护，为大数据分析奠定基础。
- 2) 计算中心。采用实时流处理和批处理技术，实现对安全事件流关联分析及批量安全事件数据挖掘，提供技术支撑和运行平台；同时采用弹性检索组件实现对安全事件的追溯取证。
- 3) 服务中心。对计算平台和数据访问提供二次接口，满足不同安全管理、安全分析需求。另外为保障平台的安全稳定运行，需做好平台安全、运行监控、数据管理及任务调度等基础功能。

2.2 情景感知在安全大数据平台应用

在构建统一的安全大数据分析平台实现安全监控、管理、分析的数据采集统一、安全模型统一、安全流程统一、用户界面统一，能够为图2基于情景感知的主动防御体系提供威胁感知、动态防护、深度监测和智能运营提供技术支撑，具体实现过程如下所示。

2.2.1 事前情景感知

攻击发生前主动搜集外部威胁情景和内部情景信息，转换特定的安全策略应用到防护和监测过程，对内外部威胁提前预警和制定防护策略；另一方面通过搜集威胁情报和现有资产属性匹配，实时风险预警^[6-7]。

- 1) 外部情景信息。针对外部的攻击，主要通过获取威胁情报，依靠专业的安全分析团队，分析之后形成情报的处置决策，并通过网络安全设备或终端上的安全软件来执行决策，达到针对高级攻击的防范。整个过程可以通过机器的自动化执行。威胁情报一般包括信誉情报(“坏”的IP地址、URL、域名等，比如C2服务器相关信息)、攻击情报(攻击源、攻击工具、利用的漏洞、该采取的方式等)

等。可以从CERT、安全服务厂商、防病毒厂商、政府机构和安组织获取安全预警通告、漏洞通告及威胁通告等，这些都属于典型的安全威胁情报。

- 2) 内部情景信息。主要指内部异常行为的监控，内部异常行为造成的破坏是安全事故最大的来源，外部攻击者发起APT攻击，其中的部分环节Delivery、Exploitation、Installation、Command and Control (C2)、Actions on Objectives都需要通过“内部行走”才能接触到敏感数据以达到盗取或破坏的目的；同时企业内部的威胁源包括可能准备离职有恶意的内部人员、内部人员的长期慢速的信息泄露，内部攻击也可能由具备内部访问权限的合作伙伴或者第三方发起。根据上述分析，设计不同的情景，通过获取样本，建立正常行为模型，然后分析内部网络流量或终端服务器上的行为数据来及时发现异常。内部情景主要指“主体”到“客体”的访问行为情景。主体是人或应用，客体是应用或数据。情景在这里包含的因素有5W (Who、What、To-What、When、Where) 等。常见的异常情景包括：登录异常行为中的异常时间、频繁登录失败等；业务违规行为中的高频业务访问和业务绕行等。

2.2.2 事中动态防护与监测

一方面具有已知情景进行威胁行为模式匹配，另一方面能基于网络交互上下文动态学习(自学习)和感知网络特定上下文，建立相应情景模型，对异常行为进行告警和阻断^[8]。

- 1) 威胁模式匹配。对于外部威胁情报中信誉情报防护设备直接根据安全事件某个特征进行模式匹配即可，如检测有外向连接到已知的危险实体，或者是检测作为潜在攻击线索的事件和行为特征的序列等^[9-10]。对于攻击情报中攻击源、攻击工具、利用的漏洞和采取的方式等信息进行威胁情报建模，对安全事件进行特征提取，形成特征序列和威胁情报模型的关联分析。在威胁模式匹配中，威胁情报标准及其可机读性是关键，通常采用XML等脚本语言使计算机自动识别其表述行为。目前成熟国外威胁情报标准包括CyboX (网络可观察表达式)、STIX (结构化威胁信息表达式)和Taxii (可信自动交换指标信息)等。

- 2) 威胁行为学习。威胁行为即指不符合业务

逻辑的行为，可以包括针对性的威胁行为（黑名单）和违反正常行为（白名单）。威胁行为通常表现为基于时间序列或基于统计的行为。威胁行为学习通过分析事件间的关系，建立不同事件关系序列库，利用过滤、关联及聚合等技术，从简单事件逐步生成高级事件。行为学习模型可通过业务专家或安全专家直接建立，也可通过机器学习方式加以人工知识进行建模。机器建模包括训练和验证，一般模型的准确率低于专家建模。

2.2.3 事后溯源和研判

对攻击行为进行研判和取证，并反馈攻击行为给情景感知模型以及修订防护策略，适当时候通过技术或者法律手段对入侵者进行反制。

1) 安全事件研判。采用可视化手段对可疑安全事件或预测性安全事件进行分析，确定攻击的准确性以及所造成的损失。事件研判需要专业知识性强的安全运维人员，除了具备安全知识外，对应用系统的行为也要非常熟悉。交互可视化可以大幅度提高事件研判的效率，但定制化因素较多，当然也可借用成熟的工具进行事件研判，如安全沙箱、网络分析软件等。事件研判的结果应该立刻进入处置流程，使安全事件闭环管理，从而进一步完善主动防御体系。

2) 攻击溯源取证。溯源通常是指寻找网络事件发起者相关信息，通常用在网络攻击时对攻击者的查找和定位。溯源相关的事件可以是应用层事件（应用层溯源，即查找业务的使用者，例如查特权账号使用者），也可以是网络层事件（网络层溯源，即查找特定网络报文的发送者）。溯源通过事件或者日志链层层查找，而取证则需要对系统或网络攻击行为进行还原，涉及的技术比较广，也是目前针对网络犯罪研究的热点。

3 结束语

随着威胁情报标准制定以及大数据实时流处理、机器学习的技术应用，未来网络安全防御体系将更加看重网络安全的监测和响应能力，将充分利用网络态势感知、大数据分析及预测技术，大幅提高安全事件监测预警以及快速响应能力，应对大量未知安全威胁。同时实时动态感知威胁情报、实时威胁情景学习与预测将使得安全防护

措施识别攻击成功率和精准度进一步提升，促进主动防御体系进一步成熟。

参考文献

[1] 贾约迪. 网络空间态势感知问题与研究[M]. 余健, 游凌, 樊龙飞, 周德川, 译. 北京: 国防工业出版社, 2014.

[2] 王慧强, 赖积保, 朱亮, 等. 网络态势感知系统研究综述[J]. 计算机科学, 2006(10).

[3] 陈秀真, 郑庆华, 管晓宏, 等. 网络化系统安全态势评估的研究[J]. 西安交通大学学报, 2004(4).

[4] 李蕊, 李仁发. 上下文感知计算及系统框架综述[J]. 计算机研究与发展, 2007(2).

[5] 张屹, 张帆, 程明凤, 等. 泛在学习环境下基于情境感知的学习资源检索模型构建[J]. 中国电化教育, 2010(6).

[6] 贾焰, 王晓伟, 韩伟红, 等. YHSSAS: 面向大规模网络的安全态势感知系统[J]. 计算机科学, 2011(2).

[7] 温辉, 徐开勇, 赵彬, 等. 网络安全事件关联分析及主动响应机制的研究[J]. 计算机应用与软件, 2010(4).

[8] 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(3).

[9] Serena H Chen, Anthony J Jakeman, John P Norton. Artificial intelligence techniques: an introduction to their use for modelling environmental systems[J]. Mathematics and Computers in Simulation, 2008.

[10] 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(3).

作者简介

杨维永 (1978-), 男, 研究生, 高级工程师, 主要研究方向为信息安全。

王红凯 (1984-), 男, 高级工程师, 主要研究方向为网络与信息系统安全运行维护。

靳丹 (1975-), 男, 硕士, 高级工程师, 主要研究方向为电力系统安全稳定分析与控制。

刘昀 (1980-), 女, 博士, 工程师, 主要研究方向为信息安全管理。

(收稿日期: 2015-12-10) EA
(编辑: 崔文静)