



基于大数据的能源集团统一运行监测与安全预警平台

王静¹, 高昆仑², 卞超轶³, 梁潇²

(1. 国家电力投资集团有限公司, 北京 100033;

2. 全球能源互联网研究院有限公司, 北京 102209;

3. 北京邮电大学, 北京 100876)

摘要: 能源集团信息化资产近年来急速增长, 为有效应对信息安全风险, 通过大数据和多源告警数据交叉确认机制等技术, 设计了集团型企业中对信息化设备和应用系统进行集中运行监测并实现安全预警和日志审计的平台。提出了通过对全网海量多源异构数据的全局分析, 实现信息网络安全风险与态势的实时感知。该平台对于提升信息系统运行和安全管理水平、及时预警信息风险向电力系统传播扩散等方面具有重要的研究意义和应用价值。

关键词: 大数据; 多源告警数据交叉确认机制; 多源异构数据; 态势感知

中图分类号: TP309.1

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018131

Unified operation monitoring and security warning platform based on big data in energy group

WANG Jing¹, GAO Kunlun², BIAN Chaoyi³, LIANG Xiao²

1. State Power Investment Co., Ltd., Beijing 100033, China

2. Global Energy Interconnection Research Institute Co., Ltd., Beijing 102209, China

3. Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: With the increasing of information assets of energy group, in order to effectively cope with information security risks, using the big data and the intersecting confirmation mechanics of multi-source data, a monitoring and security warning platform was designed for information devices and application systems in energy group. By the global analysis of big multi-source data, the real-time awareness of risk and situation could be achieved, which was valuable for the progress of operation and management level of information system and timely warning of the spread of information risk to the power system.

Key words: big data, intersecting confirmation mechanics of multi-source data, multi-source heterogeneous data, situation awareness

收稿日期: 2017-09-08; 修回日期: 2018-03-08

基金项目: 国家自然科学基金资助项目 (No.61471328)

Foundation Item: The National Natural Science Foundation of China (No.61471328)



1 引言

能源集团实施网络安全隔离^[1]是在当今严峻的网络安全形势下,加强全集团信息安全保护慎重决策的一项重大工程。在网络隔离建设过程中,全集团分别建设信息内网、信息外网,并配套建设统一防病毒、实现内网入侵检测、完善统一终端安全管理等信息安全综合防护措施。尤为重要的是,为了实现信息内网信息安全统一监测,逐步实现安全预警与集中审计,需要建设能源集团统一运行监测与安全预警平台。

2 技术方案

2.1 现状与需求

能源集团整体网络分为3个层次:集团总部、二级单位、三级单位,采用树形结构广域网系统实现互联。信息内网在各单位统一部署防病毒系统、终端安全管理、网络隔离与数据安全交换、内网入侵检测系统(intrusion detection system, IDS)等,总部集中部署综合办公、人力资源和采购管理等系统。随着建设范围的不断扩大,业务支撑设备和安全防范技术越来越复杂,面临如下问题。

- 难以实现各个业务平台统一的网络设备、服务器设备、安全设备、业务系统运行状态综合监控、预警和应急响应处理。
- 大量部署的安全产品的配置、管理方法都不相同,安全管理人员很难对每个安全产品都精通。
- 不同安全产品相对独立的部署方式使各个设备独立配置、各个引擎产生独立的事件报警,难以形成全局的安全风险监控,安全策略和配置难以统一协调。
- 与安全相关的数据量越来越大,难以对海量数据集中存储和分析处理,从大量、孤立的单条事件中无法准确发现全局、整体的安全威胁行为。

因此,迫切需要建设统一运行监测与安全预警平台,将业务系统运行的各项指标监控起来,将各自为战的安全产品集中监控起来,实现对安全事件和安全状态的全局管理。

2.2 技术路线

面对日益复杂的网络安全形势,传统的安全信息和事件管理(SIEM)分析难以应对高级持续性威胁(advanced persistent threat, APT)攻击^[2]等未知特征的威胁形式。统一运行监测与安全预警平台基于全网海量多源异构数据(例如告警数据、业务数据、网络数据、网管与运维数据和内控数据),通过对数据的集中分析,构建安全场景分析,实现安全风险与态势的实时感知。

平台将风险可视化技术应用到信息安全风险管理全生命周期,对事前风险合规性管理运维流程的成果进行量化,对事中发生的各类安全告警和异常行为及时感知,并将事后监测到的业务异动和事件处置运维流程情况全部汇总统一成态势感知的业务数据链,从多层面、多视角研究信息安全风险量化评估模型、态势评价模型、可视化展现框架和可视化交互技术,并加以应用。

统一运行监测与安全预警平台与传统的安全操作中心(security operations center, SOC)^[3]相比,在数据采集、大数据分析、安全态势感知、预警监控和可视化等诸多方面有重要突破和创新,主要对比见表1。

3 总体设计和应用

3.1 整体架构

平台以大数据采集和关联分析模块作为技术支撑,以风险管理、态势感知、安全预警和工单管理作为应用重点,通过公共数据服务接口集成第三方产品,最终通过可视化子系统综合呈现全网安全态势。平台总体架构如图1所示。

表 1 功能对比

功能项	统一运行监测与安全预警平台	SOC
数据采集	可实现不同厂商、不同设备、多种业务数据的集中数据采集,同时支持网络实时流量的采集	可实现不同厂商、不同设备、多种业务数据的集中数据采集
基于大数据关联分析	基于开源 Hadoop ^[4] 大数据系统进行数据存储、数据建模、内置挖掘算法分析功能	基于统计、规则的关联分析
监测预警	对未知和将来可预测的威胁 ^[5] 进行有效的管理,能对即将发生的安全事件提前告知,给出应急的处理措施,而且能够依据历史网络安全态势信息和当前网络安全态势信息,预测未来网络安全趋势	不支持
风险可视化	采用基于开源的 Echarts ^[6] 功能组件,将风险与态势统计、告警事件的趋势统计、不同安全域和业务域的风险变化情况统计等,进行灵活定制化的统计图表展示,力图将安全关注点以最为贴切用户视角的方式展示出来	全网整体态势、风险、脆弱性的展示,对应用集中监控,对不同视角的用户展示不同的内容

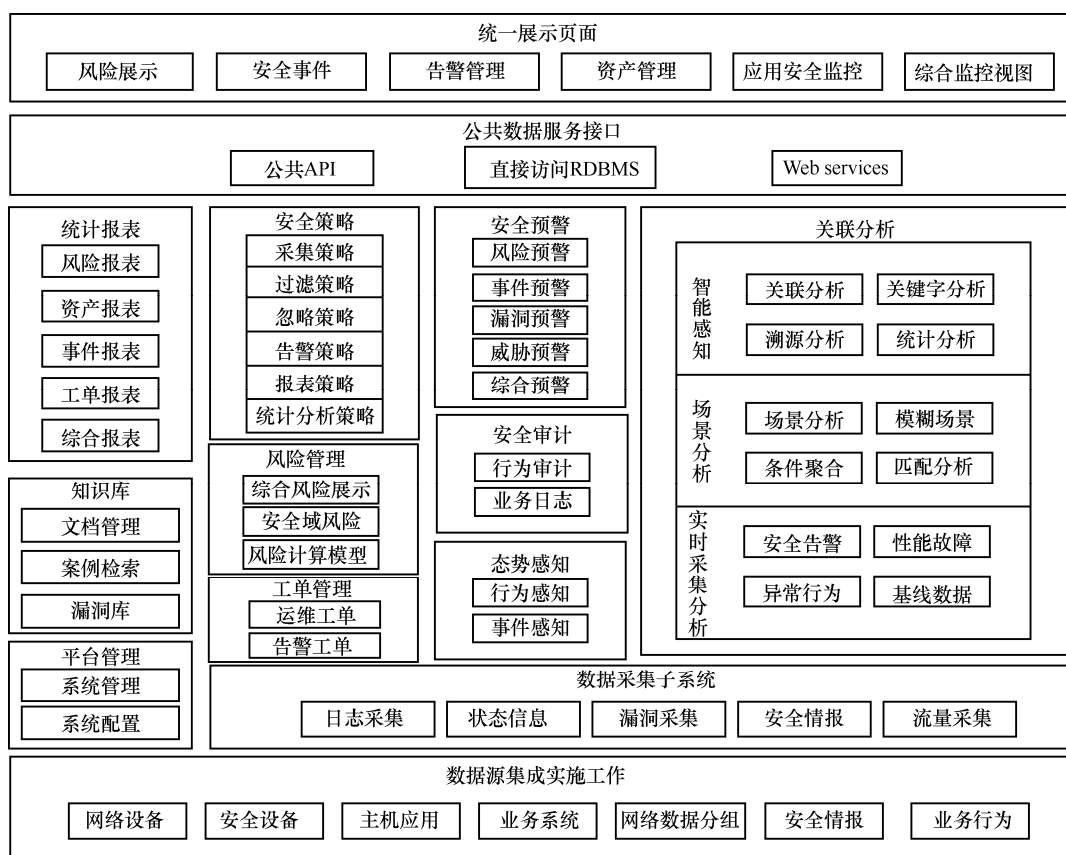


图 1 平台总体架构

3.2 功能架构

根据平台总体架构,设计基于大数据技术的统一运行监测与安全预警平台功能架构,如图 2 所示,分为日志采集层、大数据处理层、管控层、综合展现层和系统管理层。

(1) 日志采集层

日志采集层负责从网络设备、安全设备、业务系统和服务器等采集各种安全信息、日志信息、

流量信息及其他业务信息,经过数据格式标准化、数据归并、数据压缩等处理后,提交给上层数据处理平台,数据采集层提供了多种采集协议去采集各种异构数据源信息。

通过 Flume^[7]在日志系统中定制各类数据发送方,同时对数据进行简单处理,然后将数据上传到大数据处理层。来源于关系型数据库的数据通过 Sqoop^[8]上传到大数据处理层。

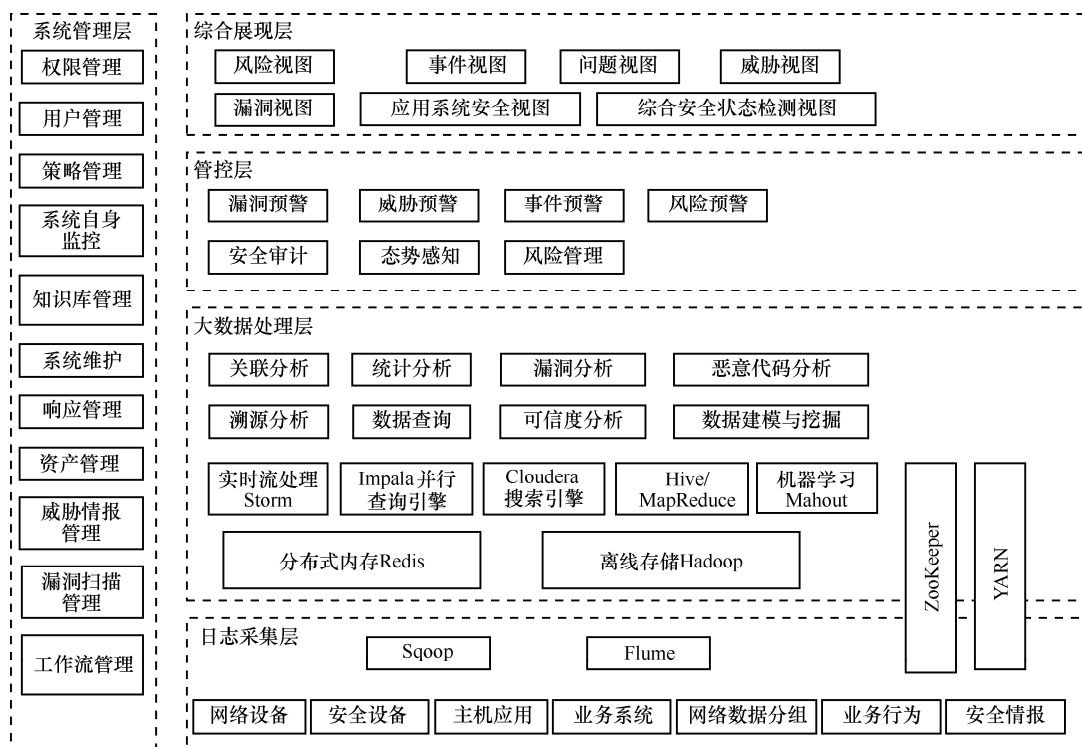


图2 平台功能架构

(2) 大数据处理层

大数据处理层的作用是对采集的数据进行预处理和存储。将需要的数据转换为结构化数据和对非结构化数据进行索引和存储。日志采集层传上来的数据分别被存储在分布式内存数据库 Redis^[9]和离线存储数据库 Hadoop 中。

实时分析包括对数据进行关联分析、统计分析、漏洞分析、可信度分析和溯源分析,结果会在综合展现层进行展示,同时会作为预警分析和风险计算的输入进行后续计算分析。

离线分析主要有历史数据统计分析、数据查询和数据建模分析,结果同样可以作为预警分析和风险计算的输入进行后续的计算分析。

(3) 管控层

管控层定制实现平台整体风险管理、预警管理和态势感知^[10]等核心预警分析功能。

风险管理是对风险进行识别和分类,对每一个风险类型进行分析,确定风险产生的因素和风险来源,并确定风险是否可以控制,对风险进行

定性和定量分析,及时对风险进行预警,提高对风险的监控效率。

预警管理利用大数据分析技术针对海量网络安全日志数据进行深层次的分析,发现数据中存在的关系和规则,根据现有的数据预测未来的发展趋势,发现潜在的安全威胁和攻击。

态势感知需根据用户实际的业务场景进行新规则的制定,根据宏观态势分析模型计算某一时间段内安全事件所属安全域或者业务系统的整体安全状态,并预测下一步整体安全走势。

(4) 综合展现层

综合展现层一方面通过丰富的图形化展示方式呈现业务网络整体安全状况,另一方面实现整个平台的灵活展示和配置管理。综合展现层提供以下 7 个概念视图,代表特定的关注领域。

- 风险视图: 主要对风险管理产生的风险和预警进行统一展示和分析,通过资产、安全域、全网多个维度进行分析和展示。
- 事件视图: 主要对实时分析产生的告警事

件进行统一展示和分析, 同样通过资产、安全域、全网多个维度进行分析和展示。

- 问题视图: 结合 IT 基础架构库运维中产生的问题, 提供统一的展示和分析。
- 威胁视图: 结合威胁情报信息和威胁预警, 进行统一的威胁展示。
- 应用系统安全视图: 从业务系统监控维度, 针对重点业务系统进行全面的安全监控展示。
- 综合安全状态检测视图: 从总体安全态势上进行全面分析, 分析整个网络环境的安全状态信息。
- 漏洞视图: 从漏洞角度对整体的漏洞信息进行综合分析和展示。

(5) 系统管理层

系统管理层主要完成系统自身监控和系统配置以及系统运维。

3.3 应用情况

统一运行监测与安全预警平台在能源集团分二期进行建设。

(1) 一期建设内容

一期建设内容主要为实现对安全设备进行统一安全日志集中管理、统一事件关联分析和统一安全事件处理流程。

对海量网络安全日志数据、外部通报和漏洞等, 匹配用户资产数据, 进行深层次的分析, 从中发现有价值的信息, 帮助用户进行安全预警。预警分析主要通过接收来自实时分析层的告警事件、风险计算层的风险预警、漏洞扫描产生的漏洞信息以及威胁情报信息, 通过分析形成有效的预警功能, 从而事前通知安全运维人员。

实现综合安全审计, 需采集用户业务系统的操作日志信息, 并配合堡垒机等行为审计产品, 进行综合关联分析, 实现业务操作全过程分析、回放和展现^[11]。

(2) 二期建设内容

平台综合各方面的安全因素, 从整体上动态反映网络安全状况, 并对网络安全的发展趋势进行预测。

实现攻击溯源过程分析, 针对安全事件进行 IP 地址回溯分析^[12]。通过源 IP 地址和目的 IP 地址的不断迭代, 将调查的轨迹逐步定位到边界。这个边界可以是网络的边界, 代表攻击者来自互联网; 也可以是内网, 代表攻击者来自内网。通过有层次的分析, 同时辅助动态的图形化展示, 可以清晰地将 IP 地址移动轨迹分析出来。

借助机器学习等算法进行自动分析处理与深度挖掘, 对网络的安全状态进行分析评价, 建立网络攻击模型, 感知网络中的异常事件与整体安全态势。

系统集中部署在集团总部, 二、三级单位部署数据采集引擎。平台在能源集团运行以来取得了良好的应用效果, 为不同视角人员(如决策者、安全运维人员、业务部门人员和系统管理等)提供不同的安全业务数据和统计分析。通过平台的管理界面将内网资产的各类信息安全事件与资产弱点进行直观展示, 加强各级人员对相关事件的重视, 并对安全事件快速响应处置, 降低安全事件误报; 将被动式信息安全管理转为主动式信息安全管理, 逐步提升对信息安全风险的精确管控、动态决策和持续改进能力。

4 核心技术

4.1 风险分析模型

平台的风险分析将 JDL (joint director of laboratory) 模型^[13]作为参考, JDL 模型将数据融合分成 5 个级别: 0 级(子对象评估)、1 级(对象评估)、2 级(态势评估)、3 级(影响评估)、4 级(过程改善)。平台的风险评估模块利用该模型进行更进一步分析, 每个功能模块在实现的同时已经为最终的风险评估提供了必要的的数据或者接口。风险分析流程如图 3 所示。

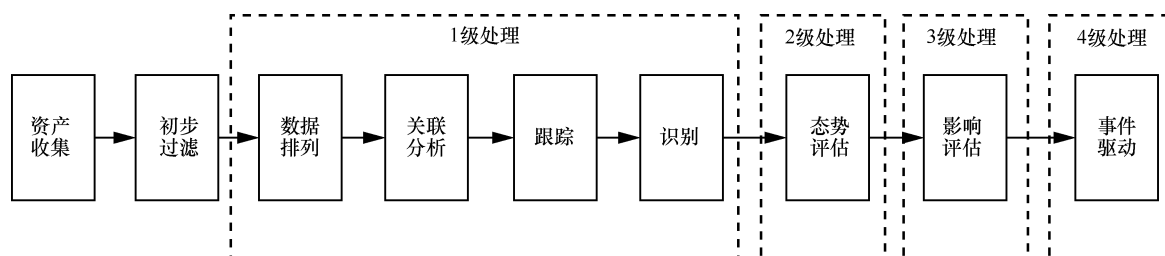


图3 处理流程

4.2 多源告警数据交叉确认机制

多源告警数据交叉确认机制主要研究模糊推理理论及算法,提高多源告警数据的可靠性。该机制的数据基础是平台全面收集的与审计相关的各种系统信息和业务信息。当多源审计结果汇总后该机制对其进行交叉验证,输出其中得到确认的审计告警信息。多源告警数据交叉确认机制如图4所示。

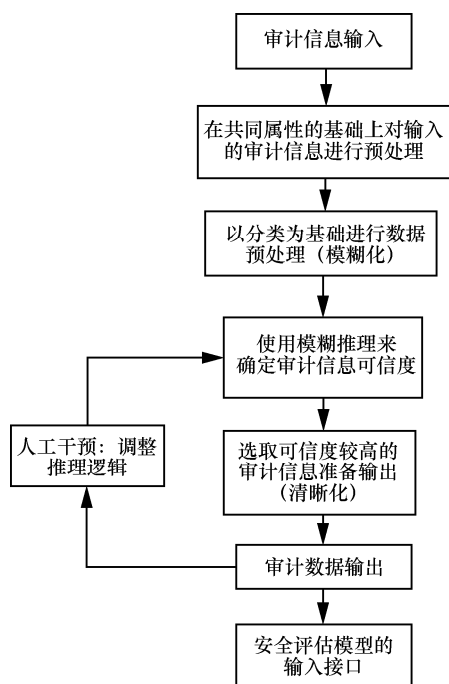


图4 多源告警数据交叉确认机制流程

4.3 安全审计模型

以真实可靠告警数据作为安全审计模型的输入,对平台监控环境的安全状态进行量化评估。技术路线如图5所示。

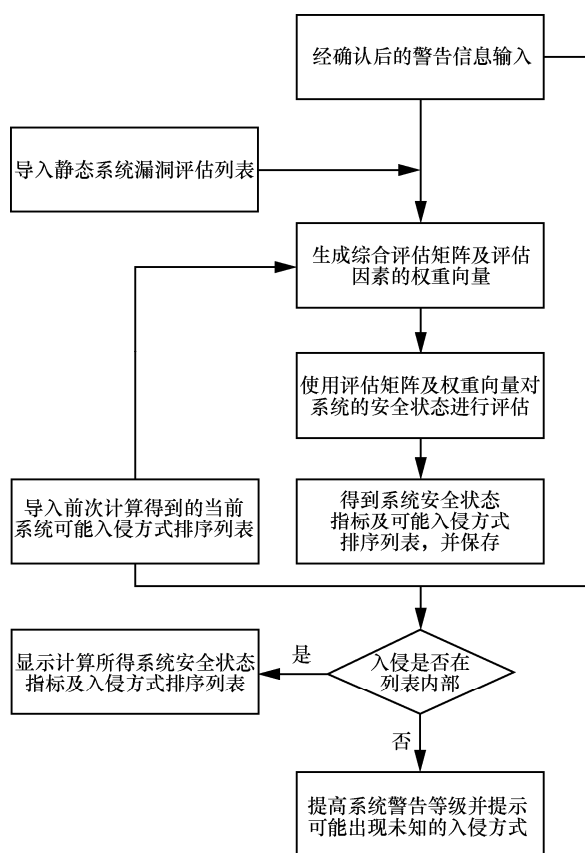


图5 技术路线

安全审计模型的计算评估过程分为以下3个过程。

- 利用系统静态漏洞评估的结果和确认后的警告信息得到综合评估矩阵,利用系统静态评估结果和前次得到的可能入侵方式排序得到评估因素的权重向量。
- 利用综合评估矩阵和评估因素权重向量计算系统当前的安全状态指标,同时可以得到系统当前可能遭受的入侵方式排序列表。

- 对前次计算所得的可能入侵方式列表进行验证。当本次的警告信息仍然被该列表包括时,则认为安全状态的评估正常,否则认为评估的结果无法对提交的警告信息进行解释。此时有可能发生新类型的入侵,应该提高安全警告的等级以引起安全管理人员的注意。

4.4 平台性能分析

本平台通过各种网络设备及服务器中采集数据,经过加工预处理后存入 Hadoop 大数据系统,再使用多种分析算法对数据进行计算处理,进而得出相关结论并提交至上层,以发出预警并进行可视化展示。为了避免或尽可能降低安全威胁造成的经济损失,识别风险、发出报警并快速响应非常重要,而平台的性能就直接影响到是否能够及时发现潜在的安全威胁。平台的性能主要受到两方面因素的影响,一是数据的采集及加工预处理过程,二是使用 Hadoop 集群对大数据的分析处理过程。将前者导致的时延记为 T_1 ,后者导致的时延记为 T_2 ,则从安全威胁产生到被发现的总时延为 T_1+T_2 。下面对这两部分时延分别进行简单分析与计算。

相对而言, T_1 的计算较为简单,可以由如下计算式给出: $T_1=\max(n/B, n/v)$ 。其中, n 表示数据量, B 表示网络传输带宽, v 表示数据预处理及导入 Hadoop 平台的速率。一般来说,数据预处理及 Hadoop 平台导入的速率在数据样式及平台导入工具(如 Flume、Sqoop)确定的情况下,主要受到平台的硬件配置影响,其中主要是预处理需要的运算资源和与存储相关的磁盘 I/O 性能。所以对于给定的数据, T_1 的大小通过对网络带宽和预处理及导入涉及的硬件资源之间的平衡来调节。

T_2 可具体分为两个部分,一部分是使用 Hadoop 集群完成对输入数据的特定分析得出结论所需时间;另一部分则是由于计算资源有限而导致的排队等待时间。对于某种特定的分析,假

定平台为其分配预定的计算资源,这些计算资源可供 N 个分析任务同时执行,若预定的资源已全部在使用中,那么后续的分析任务只能排队等待,直到前面任务完成、计算资源被释放出来后,再按先来后到的顺序继续执行。这一问题可以使用排队论模型进行分析。特别地,若分析任务的到达时间间隔服从指数分布,完成分析任务所需的时间也服从指数分布,那么该问题可看作准 $M/M/N$ (或称 $M/M/c$) 队列问题^[14],从而能够计算出平均的响应时间(分析时间与等待时间之和),即 T_2 的平均值。具体的计算式本文略去,但其中可用的 Hadoop 集群计算资源将直接影响完成分析任务所需的平均时间以及可并行的分析任务个数,从而可以通过对计算资源的控制调节 T_2 的大小。

综上所述,对于平台上的特定分析任务,可以通过理论分析计算其完成时延,并能明确网络带宽及 Hadoop 集群计算能力与其的关系,从而可以根据识别安全威胁的时延与安全威胁可能带来的经济损失之间的关系部署相关的硬件(网络、磁盘及计算节点)支持,实现经济效能的最优。

为了验证上述理论分析是否有效,开展实验记录集群的运行状况,再将统计与理论计算结果进行比对。实验采用了简单的场景设置,只关注核心的数据采集及运算处理部分,也就是理论分析涉及的内容。具体地,使用 Flume 收集日志,由 6 台服务器组成的 Hadoop 集群负责存储,其中 4 台服务器构成 Spark 集群承担计算处理任务。实验中固定了硬件资源和网络传输带宽,通过调节日志产生的速度以及变更 Spark 计算任务计划(包括数据量大小及计算复杂度、计算任务个数等)模拟不同场景,从而比较理论分析与实际测试结果的偏差。实验结果显示,理论分析计算能够很好地估算出计算时延,并指出其中的性能瓶颈所在,从而可以为硬件资源的合理配置给出相应的建议,具有非常重要的实际指导价值。



5 结束语

能源集团从信息安全风险管理视角,采用大数据技术,在物理、网络、主机、应用、数据以及进一步细化的层次上,建立起统一运行监测与安全预警平台,实现可交互的安全事件监控、综合安全审计和实时的安全态势感知,并形成一系列知识库、场景库、指标库等最佳实践成果。把关注技术细节的信息安全产品、关注事件处置流程的运维、关注整体安全态势的管理层面有机地融合在一起。因此该平台的建设不仅是安全技术手段的快速提升,同时也是管理体系上的高效改进,对能源集团网络和信息安全水平的提高起到了重要作用。

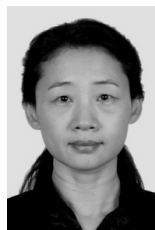
参考文献:

- [1] 王静,高昆仑,张波.基于网络隔离与安全数据交换的发电集团双网体系研究与设计[J].电信科学,2017,33(2): 163-172.
WANG J, GAO K L, ZHANG B. Research and design in dual network scheme of power corporation based on network isolation and secure data exchange[J]. Telecommunications Science, 2017, 33(2): 163-172.
- [2] 林龙成,陈波,郭向民.传统网络安全防御面临的新威胁:APT攻击[J].信息安全与技术,2013,4(3): 20-25.
LIN L C, CHEN B, GUO X M. The new threat to traditional network security defense: APT attack[J]. Information Security and Technology, 2013, 4(3): 20-25.
- [3] BIDOU R. Security operation center concepts & implementation[Z]. 2014.
- [4] Apache Software Foundation. Apache Hadoop[EB]. 2011.
- [5] 李敏,李炜,于仕,等.基于大数据分析和未知威胁感知的电网企业信息安全主动防御体系研究[J].科技广场,2016(8): 82-85.
LI M, LI W, YU S, et al. Research on information security active protection system for power grid enterprises based on big data analysis and unknown threat perception[J]. Science Mosaic, 2016(8): 82-85.
- [6] ECharts[EB]. 2013.
- [7] Apache Software Foundation. Apache Flume[EB]. 2012.
- [8] Apache Software Foundation. Apache Sqoop[EB]. 2012.
- [9] Redis Labs. Redis[EB]. 2009.
- [10] 席荣荣,云晓春,金舒原,等.网络安全态势感知研究综述[J].计算机应用,2012,32(1): 1-4.

XI R R, YUN X C, JIN S Y, et al. Research survey of network security situation awareness[J]. Journal of Computer Applications, 2012, 32(1): 1-4.

- [11] 王兴念,李宏伟,施振华,等.基于大数据的智能配电网运行监控平台关键技术研究与应用[J].电工技术,2017,2(A): 9-12.
WANG X N, LI H W, SHI Z H, et al. Research and application of key technology of intelligent distribution network operation monitoring platform based on big data[J]. Electric Engineering, 2017, 2(A): 9-12.
- [12] MURUGESAN V, SHALINIE M, NEETHIMANI N. A brief survey of IP traceback methodologies[J]. Acta Polytechnica Hungarica, 2014, 11(9): 197-216.
- [13] BLASCH E, STEINBERG A, DAS S, et al. Revisiting the JDL model for information Exploitation[C]//The 16th International Conference on Information Fusion (FUSION), July 9-12, 2013, Istanbul, Turkey. Piscataway: IEEE Press, 2013: 129-136.
- [14] 林闯. 计算机网络和计算机系统的性能评价[M]. 北京:清华大学出版社,2001.
LIN C. Performance evaluation of computer network and computer system[M]. Beijing: Tsinghua University Press, 2001.

[作者简介]



王静(1973-),女,国家电力投资集团有限公司高级工程师,主要研究方向为电力系统自动化、信息管理、网络与信息安全等。



高昆仑(1972-),男,博士,全球能源互联网研究院有限公司教授级高级工程师,主要研究方向为电力系统自动化、网络与信息安全、人工智能与大数据等。



卞超轶(1987-),男,北京邮电大学在站博士后,主要研究方向为大数据自身安全、大数据安全分析等。

梁潇(1982-),女,全球能源互联网研究院有限公司高级工程师,主要研究方向为信息安全、可信计算及大数据分析等。