Computer Science and Application 计算机科学与应用, 2019, 9(8), 1611-1624
Published Online August 2019 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2019.98181

Hans汉斯

Research and Application of IDC Security Situation Awareness System

Zongfu Li[1*], Kang Chen[1], Ang Li[2], Yang Li[1]

[1]School of Computer Science, Wuhan University, Wuhan Hubei
[2]Shenzhen Power Supply Bureau Co. Ltd., Shenzhen Guangdong
Email: *zfuli@whu.edu.cn

Received: August 7th, 2019; accepted: August 22nd, 2019; published: August 29th, 2019

Abstract

IDC equipment room is responsible for handling the massive data information at all times, and it is of great significance to discover and deal with the hidden dangers in time. In view of the lack of security awareness in data center, the lack of management work, the stubborn existence of illegal and illegal information, and the lack of trace information, this paper designs and implements an IDC security situational awareness system, which implements the early warning of abnormal device information and IP and statistics of illegal and illegal information. Firstly, the overall solution of the system is proposed. The overall architecture, logical architecture, functional modules and structural units of the system are designed. Then the specific implementation methods of each function are described. Finally, the system is fully tested and the feasibility of the system is verified, it has been put into practical use.

Keywords

IDC, Information Security, IP Detection, Keyword Filtering, Report Statistics

IDC安全态势感知系统研究与应用

李宗福[1*]，陈　康[1]，李　昂[2]，李　阳[1]

[1]武汉大学，计算机学院，湖北 武汉
[2]深圳供电局有限公司，广东 深圳
Email: *zfuli@whu.edu.cn

收稿日期：2019年8月7日；录用日期：2019年8月22日；发布日期：2019年8月29日

摘　要

IDC机房时刻承担着处理海量数据信息的重任，及时发现和处理其中的安全隐患具有十分重要的意义。

*通讯作者。

文章引用: 李宗福, 陈康, 李昂, 李阳. IDC 安全态势感知系统研究与应用[J]. 计算机科学与应用, 2019, 9(8): 1611-1624.
DOI: 10.12677/csa.2019.98181

针对数据中心安全防范意识欠缺、管理工作不到位、违法违规信息顽固存在以及日志信息无迹可寻等安全问题，本文设计并实现了一款IDC安全态势感知系统，实现了对异常设备信息和IP信息的预警以及违法违规信息的统计。首先提出了系统的整体解决方案，对系统的整体架构、逻辑架构、功能模块以及结构单元进行设计，随后描述其各个功能的具体实现方式，最后对系统进行了完备的测试工作，验证了系统的切实可行，现已投入实际应用中。

关键词

IDC，信息安全，IP检测，关键词过滤，报表统计

1. 引言

在 Internet 急速发展的推动下，互联网的服务模式和传播渠道也随之日益丰富起来。更多的企业选择 IDC 来为自己分担，将网站托管、技术应用等相关的事务全部交给 IDC 去做，以便于将更多的资源转移到自己具有核心竞争力的业务上[1]。就目前来看，我国整体的 IDC 产业突飞猛进，网络数据中心在互联网产业链所占的比例也愈发难以忽视，因其管理着海量的业务数据，安全问题逐渐成为了 IDC 绕不过的难关。然而，根据电信部门实际的统计和调查，在数据中心的对外服务环节还存在着各种问题，例如欠缺防范意识、管理工作仍不到位、违法违规信息顽固存在以及日志信息无迹可寻等缺陷。所以建设一款 IDC 安全态势感知系统具有实际必要性。

2. 系统总体设计

系统建设的目的是为了便于电信部门(管局)与下级网络数据中心(企业)的经营者们对客户的基本信息、访问日志以及网络流量情况进行日常安全协调管理工作。其中，通信管理局侧的安全管理系统(SMMS)为上级，企业侧的 ISMS 为下级，双方通过信息安全管理接口(ISMI)进行指令的上传下发、任务调度以及其他通讯工作[2]。本文研究的 IDC 安全态势感知系统属于企业侧，简称 ISMS 系统。

从系统结构上来说，本文所研究的 IDC 安全态势系统主要包括以下三个单元：控制单元(CU)、执行单元(EU)以及存储单元(DU)。每个单元各司其职，完成自己任务的同时，还需要和其他的单元产生数据以及任务交互，共同完成访问日志管理、信息安全管理以及基础数据管理等任务。具备上述功能的 ISMS 架构见图 1。

2.1. 系统单元

1) 控制单元(CU)

控制单元 CU 存在于功能区域所在的机房，任务是管理机房内的信息安全，并且利用 ISMI 与上级进行数据通信，并根据上级的要求向 SMMS 上报相应的数据反馈。同时，CU 对各个执行点的 EU 设备进行集中管理，不仅要对各类管理指令做到调度、转发以及执行等操作，还要对各类数据和流量进行分析和汇总。

控制单元 CU 的主要功能包括：基础数据的管理、访问日志的管理、信息安全的管理、报表模块的管理、系统功能的管理以及系统其他的各项管理等。而且，CU 还能够配置各种类型的结果上报以及相应

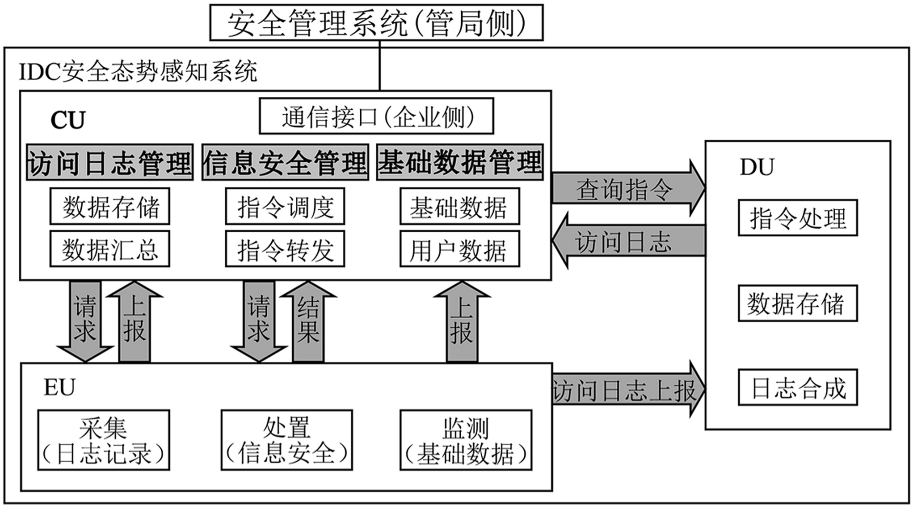

Figure 1. ISMS system architecture diagram
图 1. ISMS 系统架构图

的管理策略，并且将这些策略发送至 EU 设备。此外，CU 还可以给网站备案管理系统提供各类接口，例如 CU 和 EU 的接口(U 接口)，以便于实现各类相关信息的同步以及共享，从而给相关工作人员对系统进行管理和维护工作带来便利。

2) 执行单元(EU)

EU 的主要功能是采集网络流量数据，分析整理以及存储，最终形成访问日志。一般情况下，考虑将其布置在出口链路，对链路中的网络流量予以采集，随后依照电信部门的有关安全机制以及经营者管理条例进行初步筛选处理，产生相应的访问日志。同时，为了满足系统的其他功能需求，EU 也负责采集相关的各类数据，例如日志记录和基础信息等。对于其产生的各类访问日志以及业务内容，都需要在一定周期内根据系统需求上报至控制单元或是存储单元[3]。

在普遍的部署环境中，进入 EU 的网络流量需要再分配，也就是保证负载均衡。均衡地将网络负载分流到每一台 EU 设备上，分别进行监测等各项工作，从而降低处理延时，提高监管效率。这里的“均衡”是要求这些系统节点的承载量尽量相同，并且在某些结点出现高负载情况时进行负载迁移，重新保持节点之间的负载均衡状态。

3) 存储单元(DU)

DU 主要用于存储 EU 侧收集到的客户访问日志，对这些流量数据进行压缩、转制、合并等操作，再分类或者是统一存储。此外，DU 还提供查询功能，用来检索运营人员需要的访问日志。就数据库性能而言，应该能支持大数据的储存和检索。

2.2. 系统逻辑结构

系统存储方式采用的是分布式文件系统架构，也就是“服务器 + 硬盘”的模式。其任务是分析用户访问日志，采集、贮存、调度以及处置网络数据。系统的逻辑结构见图 2。

在大数据量的环境下，合理地利用 Hadoop 平台可以为用户提供数据查询、业务分析、设备状态监测，异常 IP 监测、流量流向分析等应用场景建模，依据上级指令，完成任务的处理工作。

海量数据查询使用 Hadoop 分布式文件系统 HDFS 和分布式列式数据库 Hbase 实现，根据常用的查询条件设计索引，实现毫秒级的数据查询响应。同时，根据多维度数据分析、统计查询等需求，系统采用分层数据处理方式，首先对数据进行清洗和过滤处理操作；然后对处理逻辑组合进行优化[4]；最后，通过调

整优化以后的客户端，进行更加灵活的查询操作。Hadoop 平台数据处理层支撑应用功能架构说明见图 3。

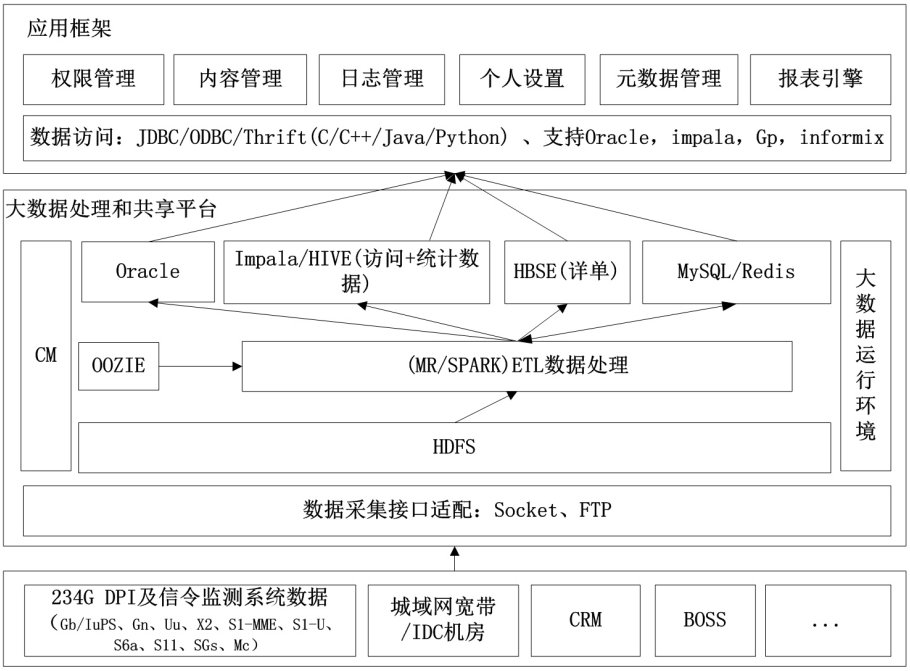

Figure 2. ISMS logical structure diagram
图 2. ISMS 逻辑结构图

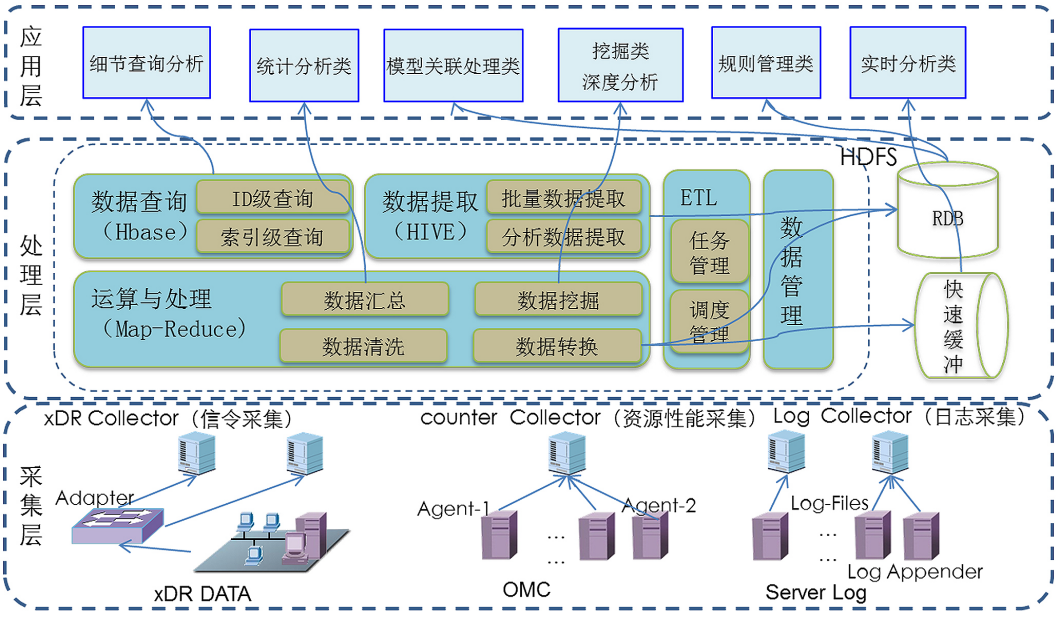

Figure 3. Hadoop platform data processing layer supports application functional architecture
图 3. Hadoop 平台数据处理层支撑应用功能架构

2.3. 系统功能结构

系统从功能架构上分类可以分为四层，自上而下层次分别是：应用层，数据处理层、通信接口层以及数据采集层(EU)。其中的前三个层次属于系统功能部分，第四部分则是 EU 数据采集设备层。系统对

应的功能结构见图 4。

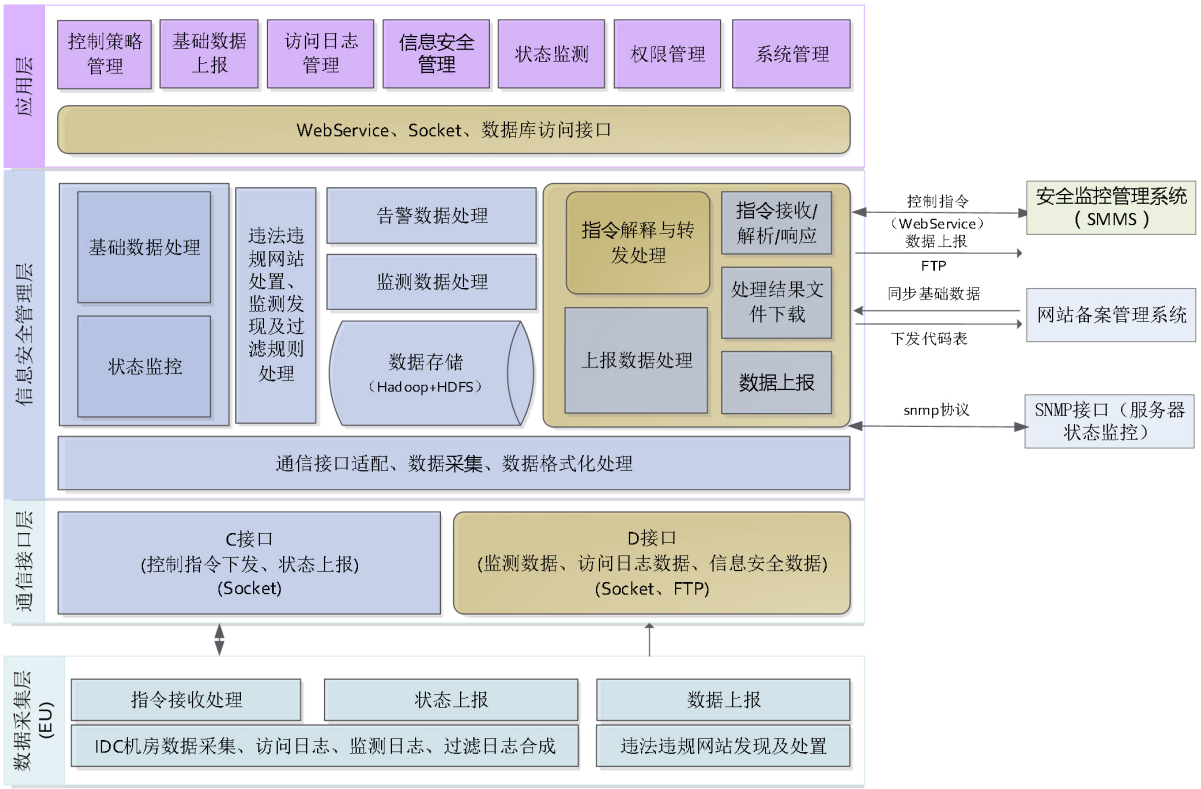

Figure 4. IDC security situational awareness system functional structure diagram
图 4. IDC 安全态势感知系统功能结构图

1) 应用层功能

应用层是 IDC 安全态势感知系统的应用部分。IDC 安全态势感知部分包含主要的三部分功能：一是完成电信部门管理数据中心基础信息的任务，同时需要采集系统设备信息，以防服务器崩溃等重大安全事故的发生；二是异常 IP 的监测以及上报；三是提供违法违规网站或流量发现规则配置以及下发，并且接受违法违规信息的监测和过滤日志的及时上报[5]。

2) 报表生成功能

系统报表是统一通过系统报表引擎来实现的，而系统报表引擎的实现基于统一语义层。语义层对底层的异构数据源进行统一封装适配(异构数据源包括 MySQL、Hadoop 等)，这使得报表引擎能够基于较稳定的数据语义层进行开发，更加专注于业务报表展现方式的灵活配置。

3) 数据处理层功能

数据处理层主要负责应用层管理策略的格式转换以及静态数据的规则匹配，并且下发给 EU 设备。同时，接收来自于 EU 设备的用户上网记录数据和 IDC 信息安全管理的数据，并且对数据进行格式化处理随后统一存储。如果遇到需要上报的，还需要对数据进行汇总、格式转化、加密以及压缩等一系列处理，之后上报给相关应用和安全监控系统。

4) 通信接口层功能

通信接口层主要负责综合数据分析与应用，可以支持系统和 EU 设备之间的策略下发、策略同步请求数据的接收以及格式化转换处理等操作。同时，还需要负责访问日志数据以及信息安全监测日志数据

的接收处理[6]。

2.4. 接口 ISMI

IDC 信息安全管理系统接口(ISMI)是 IDC 企业侧的信息安全管理系统(ISMS)与电信管理部门侧安全监管系统(SMMS)之间的接口，主要功能包括基础数据管理、访问日志管理、信息安全管理、代码表的发布等。

接口 ISMI 应该实现两种通道：命令通道和数据通道。命令通道由电信管理部门侧的系统使用，用来下发指令；数据通道由企业侧的系统使用，用来上传数据。其网络部署见图 5。

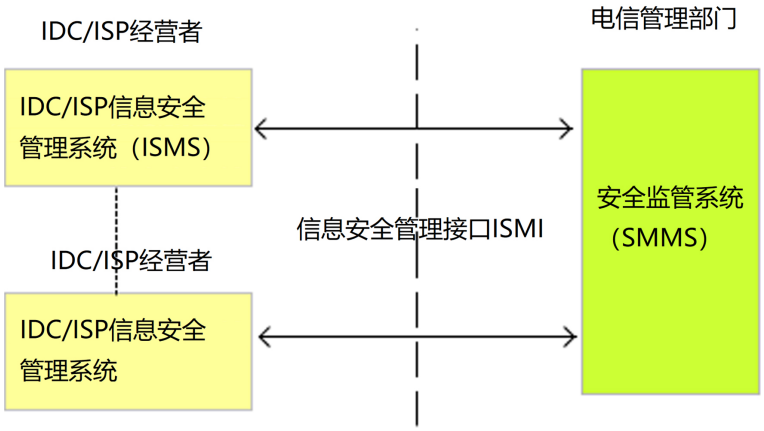

Figure 5. Network deployment diagram in ISMI
图 5. ISMI 中的网络部署图

根据以上的分析，ISMI 的设计应该实现如下几项功能，并且能够满足相应的条件部署。

1) 基础数据查询与上报接入

通过 WebService 接口接收 SMMS 下发的基础数据查询指令，并且重新将完整的基础数据通过 FTP 接口的方式上报与 SMMS，完成 ISMI 接口－基础数据通道的实现[7]。

2) 基础数据异常上报接入

通过 FTP 接口的方式向 SMMS 上报基础数据异常数据，完成 ISMI 接口－异常 IP 数据通过的实现。

3) 访问日志查询上报接入

通过 WebService 接口接收 SMMS 下发的访问日志查询指令，按照指令内容要求将访问日志整理通过 FTP 接口的方式上报与 SMMS，完成 ISMI 接口－访问日志查询上报通道的实现。

4) 信息安全记录上报接入

通过 WebService 接口接收 SMMS 下发的信息安全查询指令，按照指令内容要求将信息安全记录整理通过 FTP 接口的方式上报与 SMMS，完成 ISMI 接口－信息安全记录上报通道的实现。

5) 处理能力及可靠性接入

这里采用大叠加策略，保证大批量数据持续稳定上报，确定 ISMI 接口的稳定性，保障系统的处理能力及可靠性。

2.5. 网络部署方式

串行接入方式需要用到光保护器，将其串接在数据中心的出口链路上，原有的链路会经过光保护器进入汇聚－分流设备。这一设备可以确保系统出现设备故障时，执行快速的网络恢复操作，将业务状态

还原至事故前的网络连接，避免中断的出现。

相比于并行接入方式，串接的优点在于更为准确的控制网络流量，而且实现过程方便快捷。因为串接着光保护器，所有的流量都必须经过系统设备的处理才能进入转发流程，从而完成对流量的第一手控制，效果比并接方式更显著[8]。

但是缺点也存在同样的设备上，正因为数据链路不被复制，所以安全感知系统掌握着网络流量进出的唯一通道，这使得其存在着潜在隐患。当系统出现问题时，无论严重程度，对网络的运行的影响程度都将被放大[9]。而且，串接方案对于网络硬件设备的要求较高，可能会增加部署成本。串行接入方式网络结构见图 6。

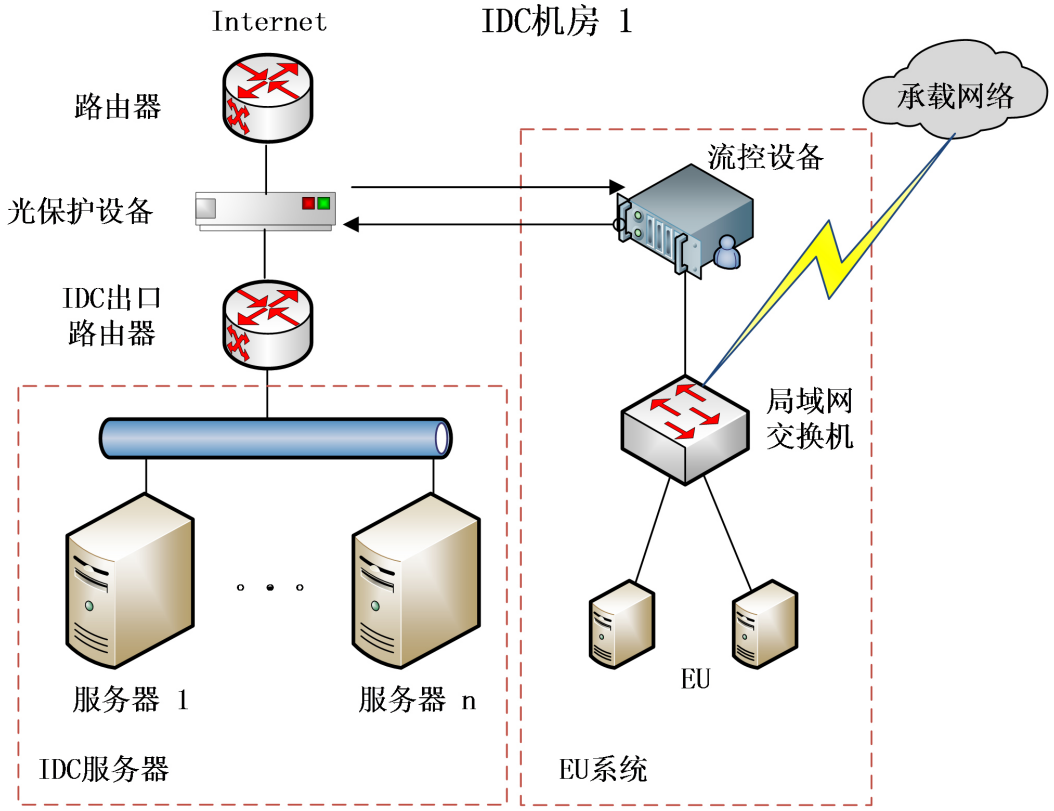

Figure 6. Serial access mode network structure diagram
图 6. 串行接入方式网络结构图

3. 系统功能实现

3.1. 设备信息采集

Sigar (System Information Gatherer And Reporter)是包含于 Hyperic-hq 产品内的一款基础包，主要用于各类设备数据的收集工作，具备良好的兼容性，支持各类平台，例如目前主流的操作系统：windows、Linux、Mac OSX 以及 AIX 等。

Sigar 还拥有多语言支持的优势，具备各类语言(如 C、C#以及 java 等)的 API，实现本功能需要用到的 API 为 java 版本，即 sigar.jar。其工作原理就是利用本地方法来调用当前操作系统对应的 API 来取得系统相关信息。例如：Linux 操作系统下的 df 指令，可以用来检查文件系统的磁盘空间占用情况；Windows 系统下的 cmd 命令 diskpart，也可以用来查看磁盘的占用情况。而 Sigar 的调用可以自动适应系统所对应

的操作系统，统一操作而且方便快捷[10]。

为了实现设备信息采集功能，需要利用这一便利的接口，采集以下两大类的系统设备信息：

第一类为静态数据，主要为：① 操作系统的各项信息。例如：内核类型、名称、版本号以及系统描述等。② CPU 信息。例如：供应商、型号以及使用者等。

第二类为动态数据，主要为：① 内存信息。例如：内存总量、当前使用量以及当前交换区使用量等。② 磁盘信息。例如：磁盘已有分区、磁盘总大小、当前已使用量、文件系统类型以及资源利用率等。③ 网络信息。例如：主机域名、当前 IP 地址、网络流量、网络接口、配置信息以及路由等。

3.2. 异常 IP 逻辑判断

首先，需要确定异常 IP 的判断逻辑以及预警机制。通过访问日志(全省 IDC 网内的上行流量)中的目的 IP 与域名两个字段与基础数据中 IP 地址段与登记域名作匹配分析出异常 IP 数据，分三种情况，逻辑如下：

1) IP 未登记

访问日志中的目的 IP 与机房信息 IP 地址段中的所有 IP 做匹配，如果访问日志中记录的目的 IP 没有在 IP 地址段登记，那么该日志信息为异常 IP 数据：IP 未登记。

2) IP 登记为保留，实际未启用

在基础数据机房信息 IP 地址段的 IP 登记信息中，IP 的使用方式分为三种：静态，动态，保留。保留的意思为该 IP 已登记但是还未启用。那么访问日志中的目的 IP 与机房信息 IP 地址段中的所有 IP 做匹配，如果访问日志中记录的目的 IP 可以与 IP 地址段内登记的 IP 成功匹配，那么就继续看该 IP 的使用状态，如果 IP 登记状态为动态或者静态，则该日志信息正常；如果该 IP 登记状态为保留，那么该日志信息为异常 IP：IP 登记为保留实际未启用[11]。

3) IP 登记，但域名有误

访问日志中的目的 IP 与机房信息 IP 地址段中的所有 IP 做匹配，如果访问日志中记录的目的 IP 可以与 IP 地址段内登记的 IP 成功匹配，并且登记状态无误，那么继续将该条记录的域名与在基础数据中该 IP 对应的域名进行比较，如果域名匹配不上，那么该日志信息为异常 IP 数据：IP 登记但域名有误。综上，判断逻辑流程见图 7。

3.3. 异常 IP 预警功能

1) IP 地址在数据库中的存储

出于实际情况的考虑，即基础数据中 IP 地址段与登记域名过于繁多，首先需要考虑节省 IP 地址的存储空间。

常规的存储数据类型为 varchar(15)，将 IP 地址作为字符数组来存储，这样做的好处是具有极高的可读性，但每个地址占用了 7~15 个字节，浪费存储空间这一缺点也十分明显。

进一步的改进是将字符串类型改成整数类型 bigint，也就是将 IP 地址 192.167.3.94 存储为整数类型 192167003094，占 8 个字节。这样虽然牺牲了数据部分的可读性，但是明显地节省了存储空间。

本文提出了进一步的优化存储方法，采用 int 数据类型，基本舍弃了数据的可读性，来保证将存储空间降到最低，仅为 4 个字节，并且有利于 IP 地址的排序，而且加快了目标 IP 的匹配和查找。以上三种数据类型的对比见表 1。

2) 匹配步骤与实现

在基础数据库构建之初，就按照 int 类型存储全部的 IP 地址，然后对整数类型的数据值进行排序。在匹配目的 IP 之前，需要调用 INET_ATON 函数完成 IP 地址的类型转换，由 string 转为 int 型。或者自定义转换函数，类似于 256 进制数的转化，带点的 IP 地址转换规则如下：

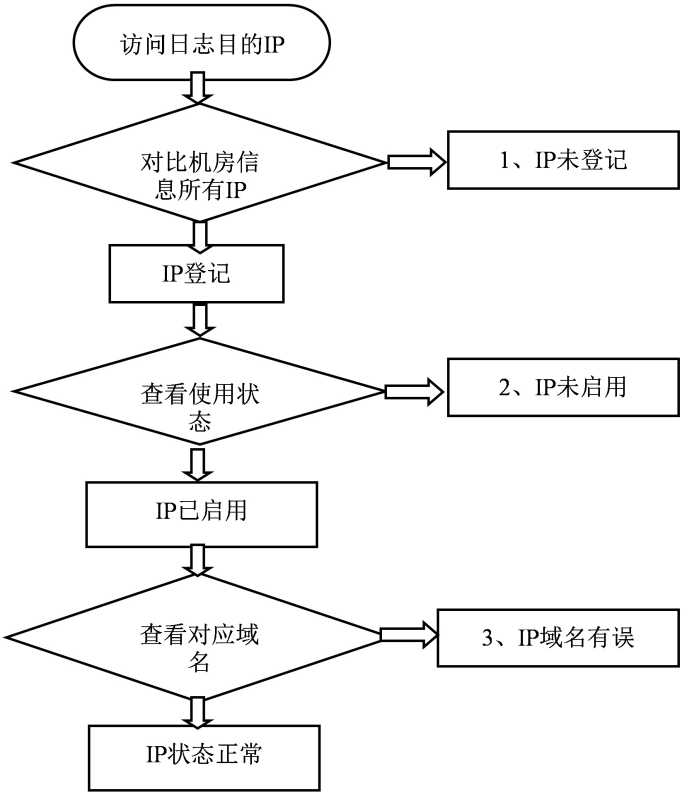

Figure 7. IP abnormal state judgment flow chart
图 7. IP 异常状态判断流程图

Table 1. IP storage type comparison
表 1. IP 存储类型对比

序号	数据类型	占用空间	存储形式	优缺点
1	Varchar (15)	7-15 字节	192.168.3.45	可读性最强占用空间较多
2	bigint	8 字节	192168003045	可读性一般占用空间一般
3	int	4 字节	3232236333	可读性最差占用空间最少

IP Number = 16777216 * w + 65536 * x + 256 * y + z

where IP Address = w.x.y.z

随后，仅需要检测目标 IP 是否存在与数据库表中即可。当转换后的 IP 地址在数据库中已基本有序时，二分逼近查找是最高效的检索方法之一，其时间复杂度仅为 O(logn)。

若不存在，则为异常 IP，异常类型为“IP 未登记”。若存在，再检测数据库中登记的状态标识码，若为保留，则为异常 IP，异常类型为“IP 未启用”；若为动态或者静态，则比较其相应的域名是否一致，来判断其属于正常 IP，或是异常类型为“IP 登记，域名有误”。

3.4. 报表引擎

报表引擎又分为四大部分，分别是：插件库、自定义报表、固定场景报表以及报表数据服务[12]。

1) 插件库

插件库是以第三方开源的展示插件为基础来实现的，其中主要的应用技术是 jQuery。系统会参考市

场使用规则形成统一的展现插件接口。对外而言，第三方展现插件处于封装状态；而对内而言，全部使用统一的展现接口。

2) 自定义报表

自定义报表归纳总结了日常的使用场景，有针对性的进行抽象操作，各类模块都是由最常用的场景抽象而来。在自定义报表里，可以自己选择布局和模块，组合成为个性化的业务报表。同时，可以将自定义报表发表在系统门户上。

3) 固定场景报表

固定场景报表是针对特定分析流程的报表场景来做设计实现的，它可以基于自定义流程的报表代码生成器，生成报表需要的基础框架代码，在生成的代码基础上进行简单程序开发，用于满足个性化报表或特定分析流程的报表需求。

4) 报表数据服务

报表数据服务是报表引擎获取业务数据的核心服务，它可以根据报表的配置信息从基础语义层获取业务数据，并以约定好的、统一的数据格式进行输出。同时，报表数据服务还负责对报表模板的配置信息进行管理。

综合上述报表引擎的结构，系统报表的功能结构见图 8。

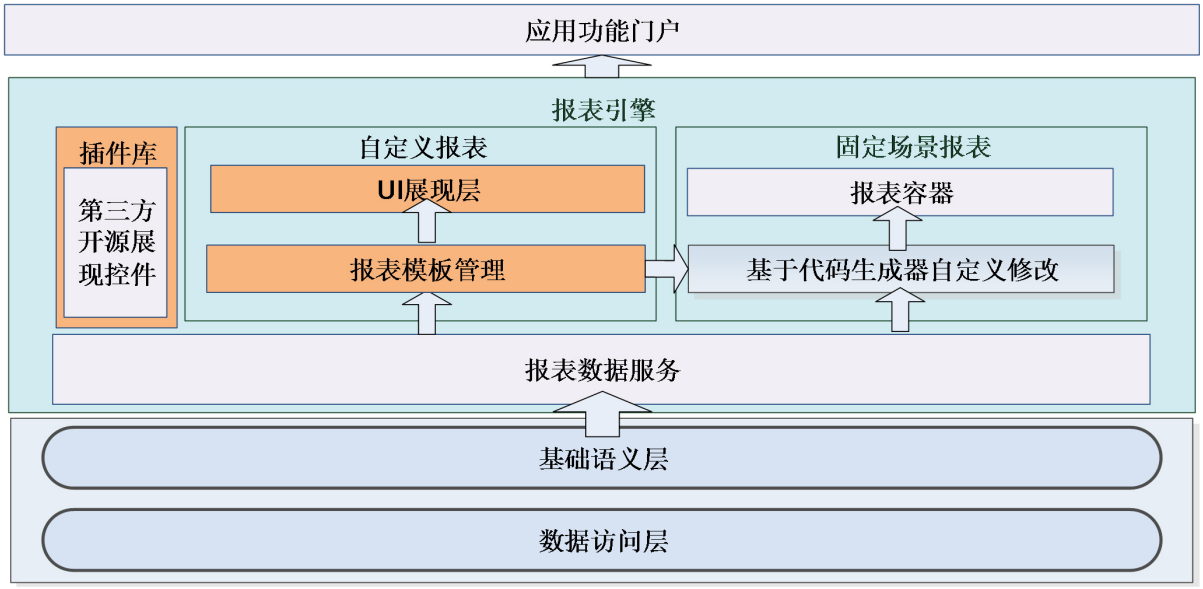

Figure 8. System report function structure diagram
图 8. 系统报表功能结构图

3.5. 报表推送分析

报表推送在信息安全的监管中处于重要的一环，是对访问日志、异常设备信息、异常 IP 信息以及违法违规信息的总结和整理。同时，也是为了响应管局系统必须存在的一项功能，IDC 需要定期向管局侧的系统发送这样的统计报表，以便于完成电信部门的监管要求。所以，报表推送功能应该包括以下部分统计内容。

1) 设备异常统计分析，也可以称为异常机房数量统计分析，根据存储的历史数据，从时间序列分析某一指标(正常/异常机房数)随时间的变化趋势，找出其规律的方法。可以用折线图来表示指标变化或发展的趋势。

2) 异常 IP 统计分析，按全省、各个地市维度进行过滤和统计异常 IP 数据，支持趋势和占比分析，以表格形式呈现。

3) 违法违规统计分析，对于采集到的网络流量进行数据过滤并且清洗等操作，去除无关信息。随后采用 DFA 算法进行违法违规信息的统计工作。下面将举例对 DFA 算法的执行过程进行必要的讲解。

首先，需要构建敏感词库，例如加入敏感词“武汉市”以及“武汉大学”，选择最适合的数据结构，构建一棵树或者森林。这样在判断某个目标词是否属于敏感词时就容易的多，很大程度上降低了检索匹配的工作量。例如要判断目标词是否是“武汉市”，根据第一个字符，就可以确定需要检索的是哪一颗树，之后再从这棵树中匹配关联。然后需要判断这个敏感词是否已经检索完毕，这就得用到标志位来完成[13]。

这部分功能使用 HashMap 实现，仍然以“武汉市”，“武汉大学”为例，构建敏感词的树，以便于关键词的过滤。

①查询“武”是否存在于 HashMap，如果检索失败，证明这样的一棵树尚未构建，那么直接 new 这样的一棵树。转到步骤③。

②如果上一步检索成功，表示有“武”开头的敏感词，执行操作 hashMap = hashMap.get(“武”)，转到步骤①，接下来继续匹配字符“汉”，“市”。

③这时需要判断是否是敏感词的最后一个字符，如果敏感词尚未终了，就定义标志位 isEnd = 0，不然定义 isEnd = 1。

完成统计操作后，进行汇总工作。按全省、各个地市维度进行过滤和统计违法违规数据，支持趋势和占比分析，以折线图或者饼状图形式呈现。

4. 系统测试

在系统的设计与实现完成之后，需要对系统进行一系列的测试工作，来确保系统的准确性和正确性。本文主要采用的是功能测试，性能测试以及安全性测试。

4.1. 功能测试

利用手工测试的办法对测试用例中的各项主要功能逐个查看，查看是否全部的按钮和链接都已经实现。使用等价类的测试方法，分别输入有效等价类和无效等价类，查看运行结果是否正常。主要测试重点如下。

1) 基础数据的本地管理

包括如基础数据的本地管理、查询和导入核验等相关功能。具体细项如下：IDC 经营单位信息的本地添加、IDC 业务客户数据的本地添加、IDC 业务资源数据的本地添加、IDC 经营单位信息的本地修改、IDC 业务客户数据的本地修改、IDC 业务源数据的本地修改、IDC 经营单位信息的本地删除、IDC 业务客户数据的本地删除、IDC 业务资源数据的本地删除、基础数据的本地导入。

2) 信息安全管理

包括活跃资源监测，违法违规网站管理，违法信息监测、过滤，规则优先级验证等相关功能。具体细项如下：活跃资源监测功能、活跃资源监测记录内容、违法违规网站监测功能、违法违规网站监测记录内容、违法违规网站处置功能、违法违规网站处置记录内容、违法违规网站列表管理、违法信息监测发现功能(IP + 端口)、违法信息监测发现功能(域名)、违法信息监测发现功能(URL)、违法信息监测发现功能(关键词)、违法信息监测记录内容、违法信息过滤处置功能(IP + 端口)、违法信息过滤处置功能(域名)、违法信息过滤处置功能(URL)、违法信息过滤处置功能(关键词)、违法信息过滤记录内容、规则优先

级与冲突校验。功能测试项目见表 2。

Table 2. Functional test project
表 2. 功能测试项目

序号	功能名称	功能描述	输入信息	输出信息	测试问题	测试结果
1	登录	用户登录	用户名密码	登陆成功	无	√
2	退出登录	退出登录状态	退出登录	退出成功	无	√
3	查看基础信息	查看机房信息	无	机房信息	无	√
4	查看基础信息	查看设备状态信息	无	设备状态信息	无	√
5	查看预警信息	查看设备预警信息	无	设备预警信息	无	√
6	查看预警信息	查看异常 IP 预警信息	无	异常 IP 预警信息	无	√
7	查看报表统计分析	查看异常 IP 统计表	无	异常 IP 统计表	无	√
8	查看报表统计分析	查看违法违规统计表	无	违法违规统计表	无	√
9	ISMI 接口通信	利用 ISMI 与上级通信	上传指令	上传指令成功	无	√
10	ISMI 接口通信	接收上级指令	无	接收指令成功	无	√

4.2. 性能测试

本次性能测试首先模拟了实际应用中的软硬件环境以及常规的用户操作过程，尽可能地保证测试状态下的系统负荷与实际运行的情况下相同，查看此时的系统运行状态，一般用来参考的参数为响应时间，CPU，内存以及磁盘的占用率等。随后，通过延长单次系统运行时间，增加系统数据处理量等操作，再次查看系统的运行状态，此时参考的参数保持不变，以此来判断系统的可靠性以及性能状态。

主要的性能指标包括：规则容量、规则匹配准确率、数据更新时间、日志查询响应时间、时钟同步、数据备份与恢复等相关指标。

具体细项如下：信息安全规则容量、信息安全规则匹配准确率、信息安全记录查询响应时间、基础数据更新时间、基础数据查询响应时间、基础数据异常监测错漏率、访问日志记录入库时间、访问日志查询响应时间、访问日志记录错漏率、活跃资源监测记录错漏率、时钟同步测试、数据备份与恢复测试。

采用二分逼近法进行 IP 匹配具有较高的性能优势，在面对百万级的数据量时，匹配时间在 0.1~0.9 秒之间；即使面对千万级的数据量，匹配时间也可以控制在 3.5 秒以内，符合系统的性能要求。

当然，匹配时间也根据匹配结果是否成功而变化，所以在测试时，需要考虑各类结果输入测试用例。测试结果见表 3。

Table 3. Abnormal IP match test result
表 3. 异常 IP 匹配测试结果

序号	数据来源	数据量(万 IP)	耗时(秒)	查询速率(万次/秒)
1	qqzeng-ip.dat	1836.1836	3.213	571.4857143
2	qqzeng-ip.dat	694.06940	1.215	571.250535
3	qqzeng-ip.dat	1672.1672	2.91	574.6279038
4	qqzeng-ip.dat	811.08110	1.42	571.1838732
5	qqzeng-ip.dat	1570.1570	2.728	575.5707478

Continued

6	qqzeng-ip.dat	1534.1534	2.679	572.6589772
7	qqzeng-ip.dat	1758.1758	3.073	572.1366092
8	qqzeng-ip.dat	556.05560	0.976	569.7290984
9	qqzeng-ip.dat	665.06650	1.616	572.8396926
10	qqzeng-ip.dat	1185.1185	2.062	574.7422405
11	qqzeng-ip.dat	347.03470	0.612	569.0501634
12	qqzeng-ip.dat	206.02060	0.367	561.3640327
13	qqzeng-ip.dat	1044.1044	1.824	572.4256579

测试环境：CPU i7-7700K + DDRC400 16 G + winlO X64。

4.3. 安全性测试

安全性测试需要验证用户类型是否对应其使用权限，能否访问权限外的信息。测试对象为具备不同用户类型的用户，参考其对应的权限执行操作，不仅要求可以访问权限内的内容，还要求禁止访问其他与权限不对应的信息，在访问系统前对用户身份进行校验来实现访问控制操作。而且，用户在一段时间间隔内不对系统进行任何操作时，应该撤销其登录状态，不能再对系统执行任何操作，并且需要重新登录。安全性测试项目见表 4。

Table 4. Safety test project
表 4. 安全性测试项目

序号	操作名称	操作描述	输入信息	输出信息	测试问题	测试结果
1	登录	用户登录	用户名密码	登陆成功	无	√
2	退出登录	退出登录状态	退出登录	退出成功	无	√
3	本权限访问	访问本权限内的内容	访问请求	对应内容	无	√
4	跨权限访问	访问没有权限的内容	访问请求	禁止访问提示	无	√
5	闲置系统	长时间不对系统执行任何操作	无	登录过期提示	无	√
6	入侵访问	不执行登录操作访问系统	SQL 注入信息	禁止访问提示	无	√

4.4. 测试结果

1) 软件功能

IDC 安全态势感知系统功能需求中的基本功能已经全部实现，有部分功能还可以继续优化，进一步修改和完善。测试结果满足功能要求。

2) 软件安全性

已经从 SQL 注入、身份验证、用户权限访问以及系统超时身份认证等几个方面对系统进行了安全性和访问控制的测试，测试结果满足安全性需求。

3) 软件容错性

已经从数据库内数据的容错性进行验证，测试了超出规定数据类型的数据输入，超出规定数据范围的数据输入。系统界面的容错性，测试了非正常操作，频繁点击操作，非规定输入操作等。测试结果满足设计需求。

4) 软件性能

根据相应性能的策略完成了 IDC 安全态势感知系统性能测试。在测试的软件、硬件以及相关的配置环境下，该软件基本已经达到了预期性能指标和设计目标。测试结果满足软件性能要求。

5. 结束语

目前网络数据安全面临众多威胁，IDC 的安全管理任重道远，本文利用 Hadoop 集成框架，完成各类大数据的流量管理任务，包括设备信息采集，异常 IP 预警，报表统计分析以及 ISMI 接口等功能，在数据存储和数据匹配方面加以改进和优化，提高了系统的运行效率。该系统已投入运行，效果良好。

参考文献

[1] 唐建军. IDC/ISP 信息安全管理系统融合建设研究及应用[J]. 中国新通信, 2017, 19(19): 118-119.

[2] 胡海波. 国外 IDC 信息安全管理形成标准我国相关技术评测手段已不断完善[J]. 世界电信, 2013, 26(4): 57-60.

[3] 陈嘉宁. 基于主成分分析和蚁群优化方法对 IP 流进行网络异常检测[J]. 计算机测量与控制, 2018, 26(5): 188-192.

[4] 吴小花, 何晓报, 赵东卓. 基于频繁 IP 地址异常流检测[J]. 长春工业大学学报(自然科学版), 2012, 33(6): 625-628.

[5] 李振国, 郑惠中. 网络流量采集方法研究综述[J]. 吉林大学学报(信息科学版), 2014, 32(1): 70-75.

[6] Wan, D. and Cao, M.H. (2009) Information Security Management System of IDC Based on ISO27001. Information Security & Communications Privacy.

[7] Wang, H.Y. (2015) Information Security Services: Cornerstone of Network Operation in the M-ICT Era. *ZTE Technologies*, No. 6, 23-24.

[8] 杨志文, 刘波. 基于 Hadoop 平台协同过滤推荐算法[J]. 计算机系统应用, 2013, 22(7): 108-112.

[9] 黄德才, 陈欢. Hadoop 平台下海量数据排行榜过滤算法[J]. 计算机系统应用, 2012, 21(3): 111-115.

[10] Riyaz, P.A. and Varghese, S.M. (2016) A Scalable Product Recommendations Using Collaborative Filtering in Hadoop for Bigdata. *Procedia Technology*, **24**, 1393-1399. https://doi.org/10.1016/j.protcy.2016.05.159

[11] Cai, L., Guan, X., Chi, P., *et al*. (2015) Big Data Visualization Collaborative Filtering Algorithm Based on RHadoop. *International Journal of Distributed Sensor Networks*, **11**, 1-9. https://doi.org/10.1155/2015/271253

[12] Liu, S.Q. and Min, Q.I. (2016) Research and Implementation of Hadoop-Based Social Big Data Collaborative Filtering Personalized Recommendation. *Modern Computer*, **32**, 74-78.

[13] 潘富斌. 基于 Hadoop 的安全云存储系统研究与实现[D]: [硕士学位论文]. 成都: 电子科技大学, 2013.