



安防大数据应用研究

杨德利, 张 涛

(中国电信股份有限公司广州研究院 广州 510630)

摘 要:描述了安防大数据的几个来源,论述了安防大数据实施中存在的难题,并据此提出了安防大数据实施的要点和步骤,最后重点分析了安防大数据的整体架构、与现有视频监控平台的关系以及基于 GPU 的安防大数据平台的系统架构。安防大数据实施时必须根据数据特点及算法成熟度分阶段进行。

关键词:大数据;安防;GPU

doi: 10.11959/j.issn.1000-0801.2015178

Research on Big Data Application for Security

Yang Deli, Zhang Tao

(Guangzhou Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China)

Abstract: Several data sources of big data on security were described, a closer look was taken at the implementation of big data platform on security, the main points and steps during the implementation were also proposed. Specially, detail analysis on the overall architecture and system structure of the big data platform on security were provided, the relationship with the existing video surveillance platform was also described. The implementation of big data platform on security must be promoted on the scale of data characteristics and algorithm maturity.

Key words: big data, security, graphics processing unit

1 引言

经过十几年的高速推进,平安城市建设过程中安装了数量庞大的监控摄像头,这些摄像头采集的视频成为了各类案件侦破的重要源头,为城市的长治久安立下了汗马功劳。但随着社会的发展,目前对监控视频的利用也存在着越来越突出的问题。一是事后查看,监控视频由于数量巨大且没有相应的分析、告警功能,只能在需要时以人工的方式检索视频,不仅耗费了大量人力物力,而且效率非常低下;二是没有对监控视频的内容进行分析和挖掘,无法

自动识别视频所包含的事件信息;三是视频信息只限于平安城市的单一作用,没有挖掘这些视频的丰富信息,供智慧城市、智能交通等相关行业和领域使用。

目前,各地智慧城市的建设正如火如荼地开展。智慧城市是运用信息和通信技术手段感测、分析、整合城市运行核心系统的各项关键信息,从而对包括民生、环保、公共安全、城市服务、工商业活动在内的各种需求做出智能响应。在智慧城市中,数据来源和获取这些数据源的方式是智慧城市建设的關鍵。现阶段在国内外,智慧城市主要案例包括利用传感器技术将水、电、油、气数字化并连接起

收稿日期:2015-06-04;修回日期:2015-07-10

论文引用格式:杨德利,张涛. 安防大数据应用研究. 电信科学, 2015178

Yang D L, Zhang T. Research on big data application for security. Telecommunications Science, 2015178

来;利用传感器技术的空车位导引;利用RFID、激光扫描、自动拍照等技术对车辆的识别;利用移动互联网技术的智慧医疗等。从这些案例可以看出,智慧城市还主要依靠传感器、RFID等技术采集的数据,而对于更加丰富的视频数据则利用不多,这也在很大程度上阻碍了智慧城市的发展。

图形处理器(graphics processing unit,GPU)已不仅仅用于3D处理领域,由于在浮点计算、并行计算领域具有超强的能力,现在也越来越多地用于通用计算领域,甚至被人解释为general computing unit。基于此,目前出现了很多将GPU嵌入MapReduce进而提高Hadoop并行计算能力的框架模型,如Mars、MapCG、SkePU等。因此,在对安防视频数据进行分析挖掘时,选择这些框架也会大幅提高系统的并行能力。

本文将在分析安防大数据的特点的基础上,阐述目前实施安防大数据的难题,并给出安防大数据的整体架构、流程以及实施步骤。

2 安防大数据特点

2.1 安防大数据来源

从广义上讲,安防大数据的数据源不仅仅包括平安城市建设的摄像头视频数据、卡口数据,也包括能对城市安全防范起作用的其他数据,如互联网数据、通信数据等。安防大数据具体可分为如下几类。

(1) 视频数据

这部分数据包括各地平安城市建设的大量视频监控数据,交警、城管、交委、海关、检验检疫等部门建设的视频监控数据以及接入电信运营商公共平台的视频数据等等。随着高清时代的到来,视频数据已成为最重要的安防数据,也是挖掘工作量最大的安防数据。目前,各个企业、商铺、家庭等根据各自的需要也自行建设了相对独立的安防监控摄像头,这些摄像头采集的视频数据由于难以采集、整合,暂时不会作为安防大数据去分析、处理。

(2) 卡口、电子警察数据

对于一线城市而言,卡口一般有几百到上千个,每天将产生上万条过车记录;而电子警察则有几千到上万个,这些都构成了安防大数据的主要来源。通过这些数据可以开展过车数据研判分析、套牌车核查、同行车分析、区域碰撞分析、连续违法分析、初次入城分析、落脚点分析、车辆

智能封控、布控与撤防等业务。

另外,用户日常社交数据(如用户日常上网浏览记录、消费记录、社交记录等数据)和用户通信数据(如用户位置数据、通信数据、访问数据等)虽然不属于安防数据,但可作为辅助数据,与安防视频数据、卡口/电子警察数据紧密结合,将极大提高安防大数据分析、挖掘的精度和效率。

2.2 目前实施安防大数据的难点

明确数据来源只是实施安防大数据的第一步,在实施过程中还是会出现各种各样的难题。

(1) 数据整合问题

安防大数据来源不同,分别存储于相互独立的系统中,将这些数据集中于统一的平台,是安防大数据实施的基础性工作,但这项工作实施的最大障碍是行业、部门的壁垒。即使只是公安内部的视频数据,由于各省、地市也互不相通,想采集和集中也不是一件容易的事。即使数据集中,怎样找到这些不同类型数据之间的关系,从而挖掘出有价值的信息,也是很大的难点。

(2) 数据挖掘、分析算法的成熟度问题

对于安防数据中最重要的视频数据,对其进行智能视频分析和挖掘是很困难的事情。目前,除了车牌识别、人数统计、部分简单特征识别等算法较为成熟外,对视频进行事件分析、人脸识别、摘要等技术都还没达到大规模的商用水平,这也极大地制约了安防大数据的实施。

(3) 时效性问题

安防大数据的目的之一就是要解决现有安防系统内以事后查看、分析为主的数据(特别是视频数据)应用形式,代之以事前预警、实时处理为主,这对大数据处理技术的实时性提出了很高的要求。

(4) 信息安全与用户隐私问题

安防行业,特别是公安行业对数据的安全性要求非常高,这也是造成数据的区域隔离的重要原因。同时,在利用安防大数据上,如何保护用户的隐私,也是一个非常重要的课题,目前主要采用数据脱敏的办法从一定程度上消除隐患。

3 安防大数据实施

3.1 安防大数据实施要点

按照安防大数据实施的难易程度,下面从纵横两个角度进行分析。

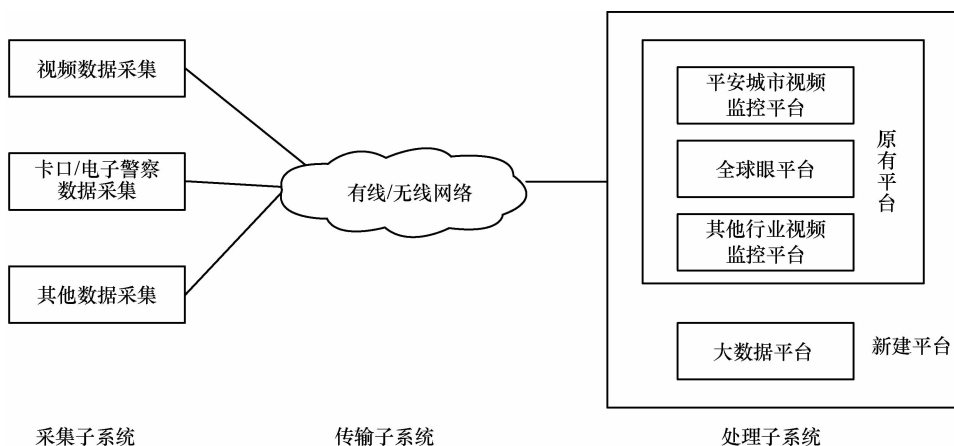


图1 安防大数据整体架构

- 纵向实施。大数据与云计算往往密不可分,在进行安防大数据的实施时,可以优先实施安防大数据的云存储,分阶段将各细分行业的数据集中存储于云平台。对于公安行业内部而言,可以先将分散的各地平安城市的视频数据进行省集中云存储,再利用云计算和大数据技术进行数据的挖掘、分析。
- 横向实施。在对数据进行挖掘、分析时,要考虑分析算法的成熟度和市场的需要。可以优先进行算法成熟高、市场需求紧迫的方向,如通过对车牌的识别实现对道路的拥堵判断等应用。

3.2 安防大数据整体架构

安防大数据整体架构如图1所示。

安防大数据系统构建于现有视频监控体系之上,包括采集子系统、传输子系统和处理子系统。其中采集子系统除了采集现有的监控摄像头采集的视频数据外,也会采集卡口/电子警察数据。传输子系统中,光纤和4G网络成为主要的传输途径。后端的处理子系统除了现有公安部门建设的平安城市的视频监控平台、运营商为政府部门、企业建设的“全球眼”平台、城管/交通委员会等部门建设的视频监控平台外,还需新建安防大数据平台。安防大数据平台从原有各平台中获取数据,并根据各部门、各行业的具体需求将分析、挖掘后的各类信息提供出去。

为了及时得到大数据平台的处理结果,提高大数据平台的实时性和价值,安防大数据平台对前端采集的数据进行实时处理,并将处理后的数据入库。

3.3 安防大数据平台系统架构

安防大数据平台的系统架构如图2所示。

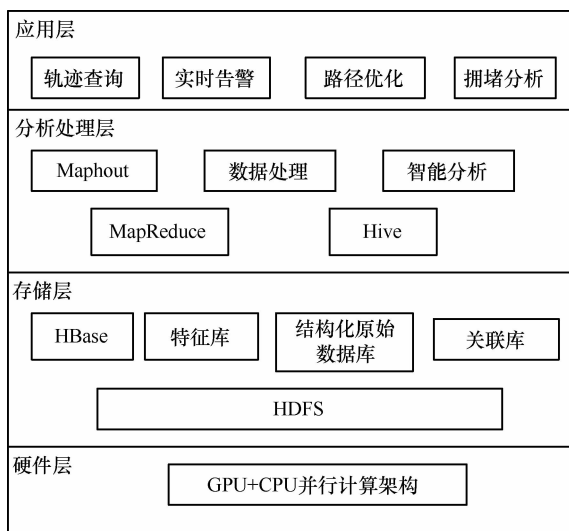


图2 安防大数据平台系统架构

从图2可以看出,安防大数据平台的系统架构共分4层,分别是硬件层、存储层、分析处理层和应用层。

硬件层采用基于GPU加速的并行计算架构,以利于将GPU嵌入Hadoop的MapReduce计算模型,为上层的分析处理打下良好的硬件基础。

存储层是构建于分布式文件系统HDFS之上的用于存储结构化、非结构化和半结构化数据的数据库,主要包括了HBase、特征库、结构化原始数据库和关联库。其中HBase作为高性能的分布式数据库,主要存储非结构化的数据;结构化原始数据库存放从前端采集的原始结构化的数据,如车辆基本信息(必要时可做脱敏处理)等;特征库主要存放从非结构化数据(如视频、图像等)中提取的特征量;关联库主要存储经分析、挖掘后的结果,这些结果可以提供给各业务直接使用。

分析处理层是大数据处理的核心部分,包括 MapReduce、Hive、Maphout、数据处理、智能分析等部分。其中,Hive 承担了数据的 ETL(extract、transform、load,提取、转化和加载)和查询功能,Maphout 提供了机器学习关键算法的实现,智能分析提供了成熟的满足业务需求的智能视频和图像的分析、识别,而 MapReduce 则通过 map 和 reduce 过程大大提高了数据处理和智能分析的性能。

应用层主要基于对安防大数据分析、处理的结果,实现满足需求的业务功能,如轨迹查询、实时告警、路径优化和拥堵分析等。

3.4 安防大数据实施关键技术

3.4.1 云存储

目前,安防行业的数据基本以 IP SAN 存储于各种互不相连的系统之中。以视频监控领域为例,监控视频数据按不同建设部门分散于不同的系统中,即使同一部门(如公安)建设的视频监控系统,视频数据也分级、分区域存放。这不仅导致了存储设备的负载不均衡、利用率低下,而且也产生了系统建设和运营成本高、能耗大等问题,云存储将能很好地解决这些问题。

表 1 是云存储和 IP SAN 存储的优劣势比较。从表 1 可知,对于以视频为主的安防大数据而言,云存储的优势是显而易见的。

在安防大数据由 IP SAN 存储向云存储演进的过程中,应结合现有系统的现状、带宽要求、急迫程度等,分阶段、分步骤实施,如可以先进行区域数据的云存储化,也可以先将某一领域(如卡口)的数据云存储化,再逐步过渡到整个安防大数据的云存储化。只有这样,才能逐步解决因部门、行业壁垒造成的数据整合难题。

3.4.2 视频智能分析

对于以视频数据为主的安防大数据而言,能不能将视频数据所包含的丰富内容分析、挖掘出来,关系到安防大数据的结果。而对视频内容特征的提取、分析和挖掘,则是视频智能分析的主要内容。

上文已述,安防大数据的云存储化需要分步骤实施,基于此,视频智能分析系统也可以分布式部署,如图 3 所示。

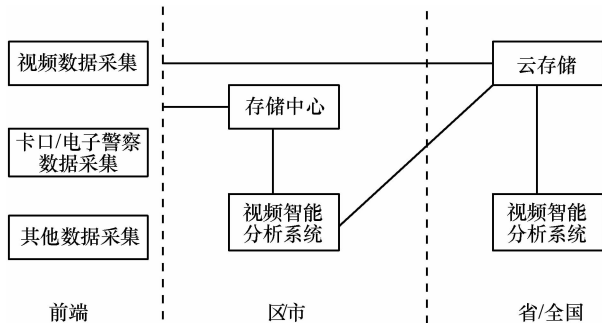


图 3 视频智能分析系统部署

从图 3 可以看出,视频智能分析系统可以部署在靠近前端的区/市分中心,也可以部署在省/全国中心。这样,部署在区/市分中心的智能分析系统可以对存储于分中心的视频数据进行分析 and 特征提取,并把结果上报至省/全国中心,而部署在省/全国中心的智能分析系统,只需对中心云存储中的数据进行分析 and 特征提取。这种分工合作的方式,使分中心传输到省/全国中心的数据量大大减少,带宽压力大大减轻,同时也可以提高视频分析的效率,满足视频大数据的实时性要求。

同时,在实施视频智能分析时,应首先从较为成熟的分析、挖掘算法(如人数统计、车牌识别等)开始,再根据各算法成熟度的发展,逐步推进其他领域的视频智能分析应用。

4 结束语

随着各行各业“互联网+”时代的到来,安防领域朝着高清化、智能化方向快速发展,安防大数据平台也被提上建设日程。在安防大数据平台建设过程中,应根据安防大数据的特点及分析挖掘算法的成熟度,选择合适的业务切入点,逐步分阶段实施。只有这样才能把安防大数据平台建设好,让安防大数据真正发挥巨大的作用。

表 1 云存储和 IPSAN 存储优劣势比较

存储方式	优势	劣势
云存储	· 适合大规模数据的存储; · 可以进行全局管理,也可以供其他业务使用; · 扩展性好,在线扩容和升级方便; · 可靠性高	· 存储规模较小时,成本较高; · 由于集中存储,对网络带宽要求较高
IP SAN 存储	· 合适小规模数据的存储; · 适合多点分布的场景	· 大规模数据存储时,成本较高; · 无法进行多点的全局管理



参考文献

- 1 列数国外智慧城市经典案例. <http://www.doc88.com/p-9005797157579.html>, 2015
Foreign smart city classic case. <http://www.doc88.com/p-9005797157579.html>, 2015
- 2 青岛成“智慧城市”双试点 首批共 20 个城市. http://blue.qingdaonews.com/content/2013-10/10/content_10035621.htm, 2013
Qingdao into a “smart city”, the first pilot has 20 cities. http://blue.qingdaonews.com/content/2013-10/10/content_10035621.htm, 2015
- 3 卢风顺, 宋君强, 银福康等. CPU/GPU 协同并行计算研究综述. 计算机科学, 2011, 38(3)
Lu F S, Song J Q, Yin F K, *et al.* Survey of CPU/GPU synergetic parallel computing. Computer Science, 2011,38(3)
- 4 Hong C, Chen D, Chen W, *et al.* MapCG: writing parallel program portable between CPU and GPU. Proceedings of the 19th International Conference on Parallel Architectures and Compilation Techniques, Vienna, Austria, 2010: 217~226
- 5 He B, Fang W, Luo Q, *et al.* Mars: a MapReduce framework on graphics processors. Proceedings of the 17th International Conference on Parallel Architectures and Compilation

Techniques, Toronto, Canada, 2008: 260~269

- 6 Enmyren J, Kessler C W. SkePU: a multi-backend skeleton programming library for multi-GPU systems. Proceedings of the 4th International Workshop on High-Level Parallel Programming and Applications, New York, NY, USA, 2010: 5~14

[作者简介]



杨德利,男,现就职于中国电信股份有限公司广州研究院,主要研究方向为视频监控、大数据。



张涛,男,现就职于中国电信股份有限公司广州研究院,主要研究方向为视频监控、大数据。

2015178-5

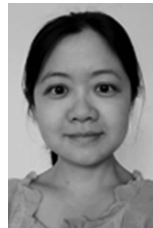
(上接 2015186-7)

- 2 George L. HBase: The Definitive Guide. Canada:O'Reilly Media, 2011
- 3 Dimiduk N. HBase in Action. Greenwich:Manning Publications, 2012
- 4 汤雅妃,魏进武,张云勇. 基于大数据的信令监测系统研究.邮电设计技术,2014(7):47~52
Tang Y F, Wei J W, Zhang Y Y. Research on big data based signaling monitoring platform. China Unicom Research Institute, 2014(7):47~52
- 5 傅巍玮,李仁发,刘钰峰等. 基于 solr 的分布式实时搜索模型研究与实现.电信科学,2011,27(11):51~56
Fu W W, Li R F, Liu Y F, *et al.* Study and implementation of distributed real-time search engine model based on solr. Telecommunications Science,2011,27(11):51~56
- 6 陈娜,张金娟,刘智琼等. 基于 Hadoop 平台的电信大数据入库及查询性能优化研究.移动通信,2014(7):58~63
Chen N, Zhang J J, Liu Z Q, *et al.* Research on loading and querying performance tuning of telecom big data based on Hadoop platform. Mobile Communications, 2014(7):58~63
- 7 张雨,蔡鑫,李爱民等. 分布式文件系统与 MPP 数据库的混搭架构在电信大数据平台中的应用.电信科学,2013,29(11):

12~16

Zhang Y, Cai X, Li A M, *et al.* Application of distribute file system & MPP database mashup architecture in telecom big data platform. Telecommunications Science, 2013, 29(11):12~16

[作者简介]



陈丽,女,广东交通职业技术学院讲师,主要研究方向为大数据技术及应用、信息系统管理。



王锐,男,中国移动通信集团广东有限公司高级工程师,主要研究方向为大数据/数据库技术及应用、网管支撑系统。

2015186-8