



基于 MapReduce 的大规模流量分析系统的设计

唐 宏

(中国电信股份有限公司广东研究院 广州 510630)

摘 要:随着用户和网络规模的快速扩大以及精细化运营需求的增加,网络流量分析系统面临的数据规模及分析深度要求都在快速发展,针对传统技术在系统扩展性、建设成本以及分析深度方面已经很难满足目前需求这一问题,提出了一种基于 MapReduce 技术的大规模流量分析系统技术方案,对数据存储、数据分析全部并行化处理,消除传统系统存在的若干瓶颈。

关键词: MapReduce; 流量分析; 大规模

doi: 10.3969/j.issn.1000-0801.2013.12.025

A Large Scale Network Traffic Analysis System Design Based on the MapReduce Platform

Tang Hong

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

Abstract: With the rapid development of network scale and the stricter requirement for operation efficiency, network traffic analysis system is facing big challenges in terms of the scale and analyzing depth. Traditional technologies cannot meet the current requirement for the scalability, cost and analyzing depth of the system. A MapReduce based traffic analysis system was presented, which could eliminate the performance bottleneck by parallelizing the storage and computation.

Key words: MapReduce, traffic analysis, large scale

1 引言

网络流量分析是网络运营分析的重要部分,能够有效反映网络整体负载及运行状态,是网络运营、规划的重要支撑系统。随着用户和网络规模的快速扩大和精细化运营需求的增加,网络流量分析系统面临的数据规模及分析深度要求都在快速发展,传统技术在系统扩展性、建设成本以及分析深度方面已经很难满足目前需求。

另一方面,大数据技术的出现和普及为解决大规模数据分析问题提供了新的机遇。大数据技术如 MapReduce 平台通过简化存储和计算模式获得平台的高度扩展性,将集群规模从数十台扩展为数千台,将处理能力从千万条扩展

到数十亿条。网络流量分析系统中计算最密集的任务是数据汇聚,分析任务大多具有松耦合特点,天然符合大数据 MapReduce 的计算模式。本文提出了如何利用 MapReduce 技术构建新型大规模流量分析系统,并且给出重要算法的 MapReduce 分解方法,使得流量分析系统能够具备接近线性的性能扩展能力。

2 流量分析系统

流量分析系统的核心任务是通过流量数据的采集、分析,了解网络的整体资源利用状况,尤其是了解网络流量的来源以及去向分布,以便对网络运营、优化、扩容做出科学的指导。流量分析系统包括宏观层面的流量分析和微



观层面的流量分析,其中宏观层面的分析着重研究网络流量的宏观统计特性,比如网络流量的来去分布、网络流量的应用构成、分组长度的分布等,一般基于 Netflow 协议实现。宏观层面的流量分析系统核心部件介绍如下。

(1)数据源

数据源是提供原始分析数据的部件,目前主要为路由器。在配置流量监控以后,路由器会对途径的流进行采样,一般是每隔若干个分组进行一次采样,然后将流概要信息(流记录(flow record))通过固定协议发送给采集器,以较低的代价为分析系统提供概要信息。发送的信息一般包含源地址、目的地址、通信两端的端口号以及协议类型。由于使用了采样技术,这种数据源一般只能做宏观层面的统计分析,无法深入一个流内部查看详细信息。

(2)采集器

采集器用于从数据源接收流信息,并按照分析系统的要求存储到指定的位置。采集器的核心是实现协议的编解码,快速提取出流记录信息。

(3)存储器

存储器用于实现流记录的存储,为数据分析提供输入。存储器需要支持海量的存储能力,以支持日益增长的流量数据。

(4)分析器

分析器实现流量数据的分析,是系统的核心部件。分析器需要支持强大的计算能力,以在短时间内实现各种流量分析任务。

微观层面的流量分析系统实现报文级的分析,能够查看报文的每个字节以了解报文所承载的内容。其总体架构与宏观层面的流量分析系统架构大体一致,只是数据源不再是路由器设备,而变为一种专门的硬件设备。这种设备一般接在链路上,将流经链路的每个报文都截获下来并存储,并提供给分析系统进行深度分析。

3 基于 MapReduce 的流量分析系统架构

传统流量分析系统一般采用特定的小规模分布式架构,针对相应的任务设计数据的存储及分析方法,没有构建在通用的分布式文件系统和分布式计算平台上,因此系统的整体扩展性相对较低,开放性也较差。本文提出了基于 MapReduce 框架的流量分析系统,利用分布式

文件系统解决大规模数据存储问题,利用分布式计算解决大规模数据计算问题。基于 MapReduce 的流量分析系统框架如下。

系统整体基于 MapReduce 平台部署运行,数据采集器在集群的多台设备上分布执行采集任务,将收集到的数据写入分布式文件系统。数据分析模块定期启动 MapReduce 分析任务,从数据目录中导入数据,执行各种分析,并将分析结果写入易于使用的关系数据库中。整个过程可以看成是生产者和消费者模型;数据采集器是生产者,将收集到的数据写入存储模块;数据分析是消费者,定期地从存储模块访问存储的数据。所有核心过程全部分布化进行,没有任何性能计算瓶颈。

(1)数据采集

数据采集任务相对简单,在一个大型网络流量分析系统下,数据源往往有数十台甚至上百台路由器,一台采集器无法支持全网的数据收集。在本系统方案中,采集器部署在多台服务器上,同时作为分布式文件系统的客户端执行写入。在一个实际的大型网络中,即便是在 5 000:1 的采样设置下,核心路由器每秒产生的流记录数达到上百万条,大致需要 10 台采集器才能够完整地接收这些流记录。

(2)数据存储

数据存储负责将采集到的记录保存到系统以供后续分析。对流量分析系统而言,流记录具有结构化属性,不考虑性能因素,将这些数据存放于 NoSQL 数据库是一个较好的方案。但对于一个大型网络,流记录速率达到百万每秒,数据库很难支持如此高的并发速率。另一种解决方法是每个采集器每隔一定的周期将采集到的流记录写入分布式文件系统,数据按照文件的方式存放。这种方案可以保证采集器有线性扩展能力,整体的 I/O 吞吐量随着采集器的增加呈线性增长。

(3)数据分析

数据分析负责提供网络流量数据的分析功能。数据分析的主要任务包括流量流向的分析、流量应用构成分析以及特定的组合分析等。流量分析每隔一定周期执行,以处理积累的一批数据。在 MapReduce 框架下,流量分析的算法需要按照 MapReduce 的模型进行并行化,具体在第 4 节阐述。

4 数据分析算法及并行化

流量分析的核心任务是弄清网络的整体负载状况,包

括进出网络的流量大小、流量出入方向、流量的应用构成等。需要分析的流量数据核心是一组记录集合,每条记录代表流经路由器的流采样信息,核心信息为:FlowRecord=<src_ip,dst_ip,ip_proto,src_port,dst_port,packets,octets,time>,每个属性介绍如下。

- src_ip:流的源地址,代表流从哪个 IP 地址产生。
- dst_ip:流的目标地址,代表流向哪个 IP 地址。
- ip_proto:流的 IP(互联网协议),代表流的协议类型,如 UDP、TCP 等。
- src_port:流的源端口号,代表 UDP 或 TCP 的源端口号。
- dst_port:流的目的端口号,代表 UDP 或 TCP 的目的端口号。
- packets:流的分组数量。
- octets:流的字节数。
- time:流的产生时间,详细地包括流开始和结束时间。

MapReduce 算法分 map 和 reduce 两个阶段,其中 map 阶段将输入数据进行转变,以<key,value>形式输出。在 reduce 阶段,map 阶段输出的具有相同 key 的记录会被组织在一起,reduce 仅需要将这些记录按照分析任务进行使用,然后按照预定的格式输出到一个指定目录。整个 MapReduce 算法的核心就是 map 和 reduce 两个函数的实现。以下是最典型的基于 MapReduce 的流量矩阵分析算法。

任务说明:流量矩阵的目标是统计出流量流向矩阵,每个矩阵元素(i,j)代表从第 i 个区域流向第 j 个区域的流量大小。每个区域可以通过 IP 地址集合表示。

(1)map 函数

输入:key=ID,value=FlowRecord=<src_ip,dst_ip,ip_proto,src_port,dst_port,packets,octets,time>

处理流程:

```
for each record
    srcRegion=LocationHelper.findRegion(record.src_ip);
    dstRegion=LocationHelper.findRegion(record.dst_ip);
    mapOutputKey=srcRegion<->dstRegion;
    Collector.write(mapOutputKey,record.octets);
end
```

(2)reduce 函数

输入:key=SrcRegion<->DstRegion, ValueList=OctectList

处理流程:

```
for each key
    sum=0;
```

```
for each value in ValueList
    sum+=value;
end
Collector.write(key,sum);
end
```

实际上,流量分析系统还有很多任务,比如流量应用构成分析等。这些任务核心就是在 map 函数上定义输出 key,在 reduce 阶段汇总,具体的 key 转换定义一般都能够根据分析目标直接获得。比如分析流量协议构成,只需要将 record.ip_proto 定义成 map 的输出 key 即可。

5 实验结果

为了验证基于 MapReduce 的 Netflow 流量分析系统的扩展性,在实验室环境下使用 8 台物理服务器,全部安装 CentOS 6.3 操作系统,然后在所有服务器上安装、部署 Hadoop 分布式环境,搭建流量分析分布式集群。执行流量流向分析 MapReduce 作业,测试如下。

(1)在集群资源保持不变情况下,逐渐增大 MapReduce 分析任务的输入数据量,统计流量流向分析的计算时间,统计结果见表 1。

表 1 可扩展性验证			
数据量	45 GB	74 GB	104 GB
处理时间	5 min 6 s	6 min 41 s	9 min 9 s

(2)在保证输入数据量不变的情况下,逐渐增加集群节点的规模,同样使用流量流向统计 MapReduce 分析算法的计算时间来验证集群的可扩展性,统计结果见表 2。

表 2 可扩展性验证			
节点数量	4 个节点	6 个节点	8 个节点
处理时间	20 min 38 s	14 min 12 s	9 min 9 s

实验表明,使用 MapReduce 进行 Netflow 流量分析,具有很好的扩展性,计算能力能够通过堆叠服务器获得接近线性的能力提升。

6 结束语

本文提出了一种基于 MapReduce 的大规模流量分析系统架构,该系统利用分布式文件系统解决大规模数据存储问题,利用分布式计算解决大规模数据分析问题。基于这种框架,大部分流量分析系统的密集计算任务都可以轻松地被分解,从而在较短的时间内完成大

(下转第 163 页)

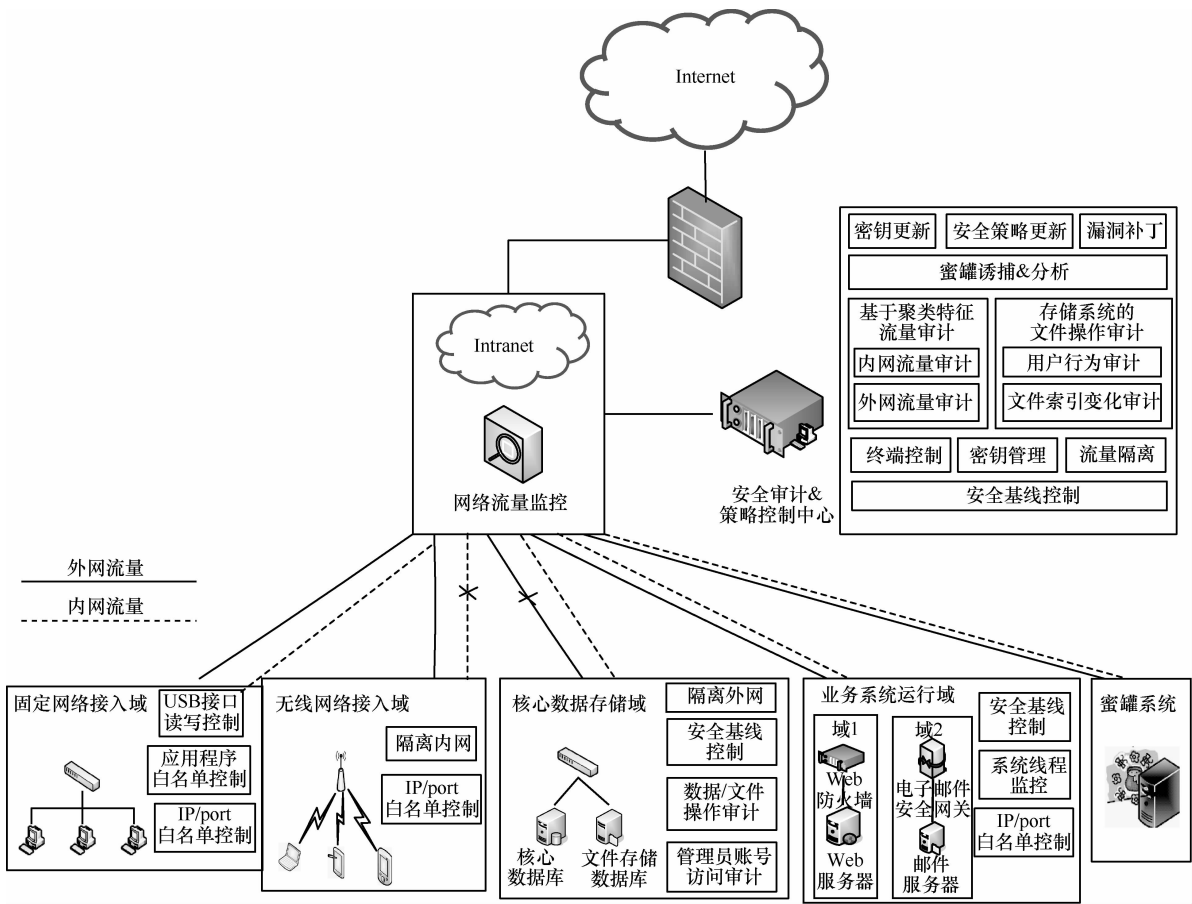
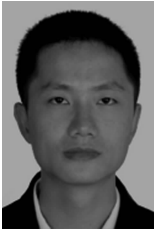


图 5 APT 攻击动态防御的部署示例

[作者简介]



刘东鑫,男,中国电信股份有限公司广东研究院工程师,主要研究方向为网络与信息安全等。

王帅,女,中国电信股份有限公司广东研究院工程师,主要研究方向为 IP 网络安全、信息安全等。

沈军,女,中国电信股份有限公司广东研究院高级工程师,主要研究方向为 IP 网络安全和网络应用安全技术等。

刘国荣,男,中国电信股份有限公司广东研究院工程师,主要研究方向为信息安全、数据库安全等。

金华敏,男,中国电信股份有限公司广东研究院高级工程师,主要研究方向为 IP 网络安全和 IT 系统安全技术等。

(收稿日期:2013-08-25)

(上接第 157 页)

规模的数据计算,解决现行流量分析系统面临的扩展性和成本问题。

参考文献

1 White T. Hadoop: the Definitive Guide. O'Reilly Media, 2011
2 Ghemawat S, Gobioff H, Leung S. The Google file system. ACM SIGOPS Operating Systems Review, 2003(5)

[作者简介]



唐宏,男,中国电信股份有限公司广东研究院高级工程师,主要研究方向为 IP 网络与云计算技术研究。

(收稿日期:2013-12-01)