doi:10.3969/j.issn.1000-436x.2012.11.011

网络管控系统的柔性可重构结构设计

李玉峰，邱菡，刘勤让，兰巨龙
（国家数字交换系统工程技术研究中心，河南 郑州 450002）

摘 要：网络管控系统是面向互联网“可管、可控、可信”需求提出的新设备形态。骨干链路级管控设备的体系结构目前基本沿袭核心路由器的设计思路，存在数据平面、控制平面分离无法满足管控系统软硬件协同深层次数据处理需求，刚性封闭系统构建方式无法满足新业务识别柔性需求的问题。由此提出网络管控系统的软硬件交换式松耦合协同处理结构和软硬件循环迭代迁移结构和方法，可实现软硬件融和式协同数据处理，满足了新业务识别的柔性迁移需求。

关键词：网络管控系统；体系结构；可重构；循环迭代；迁移
中图分类号：TP393.08 **文献标识码：**A **文章编号：**1000-436X(2012)11-0084-07

Design of network management and regulation system over flexible reconfiguration

LI Yu-feng, QIU Han, LIU Qin-rang, LAN Ju-long
(National Digital Switching System Engineering and Tech logical Research Center, Zhengzhou 450002,China)

Abstract: Network management and regulation system is a new arising appliance for the manageable, controllable and trustworthy demand of the Internet. The architecture design of the backbone link-level management and regulation system usually follows the core router, in which the separation of the data plane and the control plane could not meet the requirement of HW/SW data co-process, the rigid connection with closed building style could not satisfy the flexible identification of new applications. The loosely coupled switching HW/SW co-process architecture and the iterative looping HW/SW migration architecture were proposed, which could satisfy the above HW/SW co-process demand and the new application flexible identification demand.

Key words: network management and regulation system; architecture design; flexible reconfiguration; iterative loop; migration

1 引言

互联网逐步成为各种通信基础设施的统一平台，承载的业务越来越多，并且呈现出新的特点：少数低价值的业务正在吞噬越来越多的带宽资源，如 P2P (peer to peer)类业务流量占用了近三分之二的骨干网带宽[1]，运营商正面临着增容不增收的窘境；网络游戏、高清晰 IPTV 等各种新兴业务不断涌现，使得网络业务日趋复杂多样，对传统的粗放式网络运行维护和运营模式提出了更为严峻的挑战，由于缺乏精细化运营的基础，运营商愈发难以掌控客户的网络行为，无法进行针对性的业务开发和营销；更为严重的是，对于一些如非法宣传、网络病毒、网络攻击、垃圾邮件等不良信息，由于缺

收稿日期：2011-12-02；修回日期：2012-06-20
基金项目：国家高技术研究发展计划（“863”计划）基金资助项目（2011AA01A103）；国家重点基础研究发展技术（“973”计划）基金资助项目（2012CB315900）
Foundation Items: The National High Technology Research and Development Program of China (863 Program) (2011AA01A103); The National Basic Research Program of China (973 Program) (2012CB315900)

乏有效的识别和管控手段，致使其能够通过网络广泛传播、大规模泛滥，对网络安全和可信造成了严重威胁。为此，我国互联网提出了“可管、可控、可信”的具体要求，网络管控系统应运而生[2]。

从本质上讲，网络管控系统是一种感知网络、控制网络的设备，其涉及的关键技术主要包括全分组范围内容级搜索、全维信息判决、用户行为分析、不良信息检测、流量统计特征和主机行为特征分析等，这与以转发、交换为核心的传统路由器、交换机设备大相径庭。当前，具备一定业务识别和控制能力的是 DPI(deep packet inspection)类设备[3,4]，整体而言，该类设备是在传统路由器基础上进行了简单的网络管控功能增强，其低端产品基本沿用以软件处理为核心的“存储＋分析”结构，无法适应网络管控的高速处理要求，其高端产品也基本沿袭数据平面、控制平面分离的传统核心路由器设计思路，在面对网络管控系统深层次的软硬件协同数据处理需求时基本无能为力。

互联网上应用创新已经成为其发展常态，目前， 互联网高端 DPI 设备厂商普遍面临一个窘境：一款新设备刚推出，马上就需要针对新业务开发的新管控方法。新业务出现初期流量较小，可以通过升级软件、打补丁方法支持。此后，若该业务发展缓慢，流量未出现较大增长，则设备厂商可维持补丁形态；相反，若该业务发展迅猛，其流量在短期内出现了爆炸式增长，超过了软件处理的能力范围，导致软件补丁方法难以为继，这种情况下，设备厂商只能将软件补丁硬件化实现，完成软件处理的硬件迁移。目前，骨干链路级管控设备为应对高速业务识别和控制需求，普遍最大化地将设备功能使用定制电路实现，其系统构建本质上是一种刚性封闭方式，无法完成软件功能向电路的动态迁移。要实现迁移，设备厂商可选用的方法只能是推倒从来，开发出新一代的硬件平台并基于新平台完成定制电路升级，从而实现对新业务管控的支持。至此，设备厂商完成了一次“新业务出现—软件补丁发布—新业务流量激增—新平台推出—定制电路升级—迁移完成”的轮回。近年来，新业务的出现呈现出数量大、间隔短的特点，管控设备厂商面临的硬件迁移轮回周期愈来愈短，管控设备的生命周期正变得越来越短，新产品推出的间隔也越来越短。

针对上述问题，本文提出网络管控系统的软硬件交换式松耦合协同处理结构，支持软硬件无阻塞协同数据处理，提出软硬件循环迭代迁移结构和方法，满足新业务的“软件→硬件”动态迁移需求，从而给出一个“能力可扩展、功能可迁移、识控一体化”的网络管控系统结构。

2　软硬件交换式松耦合协同处理结构

网络管控系统结构遵循自顶向下设计思路，基本思想是：通过大容量多级交换组件完成系统前台各个硬件组件和系统后台各个软件子系统的松耦合无阻塞数据传输，实现软硬件数据平面的功能融合；通过系统内硬件构件、组件和软件子系统的柔性迭代组合，实现管控系统功能与性能的可重构；通过控制信息的交换实现所提的网络化管控理念[5]，支持“一点发现，全网联动控制”的管控方式。图 1 是本文所提的网络管控系统结构。

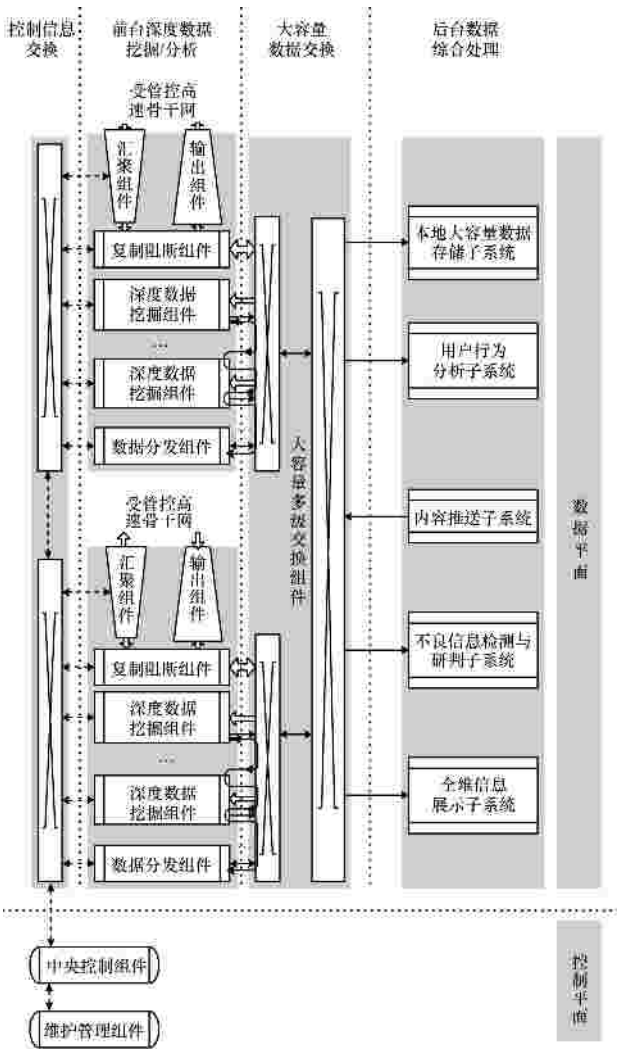

图 1　网络管控系统软硬件交换式松耦合协同处理结构

该结构中，数据平面包括线路汇聚组件、复制阻断组件、深度数据挖掘组件、数据分发组件、输出组件、大容量多级交换组件6类硬件组件以及本地大容量数据存储子系统、用户行为分析子系统、内容推送子系统、不良信息检测及研判子系统、多维信息展示子系统等软件子系统。数据平面上，硬件组件完成对报文的串行接收、封装解封装、在线业务种类识别、在线管控、串行输出和筛选操作，筛选出的数据通过交换组件送到后台软件子系统进行相应的分析处理，从而实现前台快速识别管控和后台综合处理的功能融合。控制平面包括中央控制组件和维护管理组件，完成管控系统的控制、维护和管理。受硬件工艺水平的限制，单个深度数据挖掘组件支持的规则数量和规则长度受限，为应对管控系统对规则数量和规则长度的多样化需求，该结构提供了多个深度挖掘组件的级联处理能力，数据分组输入后可通过交换组件循环交换到多个深度挖掘组件迭代处理，实现了规则数目、关键词宽度的灵活扩展，提高了系统的灵活性。

3 软硬件循环迭代迁移结构和方法

网络管控系统将网络业务划分为2大类，第1类是载荷中存在业务“指纹”的业务，针对这类业务，管控系统存在的主要技术难题是线速、深度、多模式、大容量联合“指纹”匹配。从实现层面上看，已有很多研究工作围绕该难题展开，这些工作大多都基于专用处理芯片和专用存储器件（例如三态内容寻址存储器）[6~8]，采用关键词联合匹配方法实现。第2类是载荷中无业务“指纹”或者指纹特征很难获取的业务，例如加密类网络业务、新型网络异常攻击或者暂时无法获取“指纹”特征的各类新涌现业务。针对第2类业务，管控系统面临2个技术难题：一是识别速率低。第2类业务的识别可基于分析通联过程特征实现。通联过程特征可概括为2类：流量统计特征和用户行为特征。流量统计特征中如流平均分组长、流持续时间等在加密前后变化不明显，而且新业务出现时也往往会伴随新的通联特征出现，故基于通联特征可有效识别加密、异常和新出现业务。通联特征维度高达248余种[9]，基于通联特征的业务识别具有复杂、灵活和智能特性，目前普遍基于软件实现，业务识别速率通常在兆(M)比特级。二是新业务识别“软件→硬件”迁移难。新业务出现的初期，流量较少，可以通过软件处理。随着业务的发展，其流量增大，当流量增大到超过软件处理能力时，就需要将相应的软件功能转化为硬件实现，而传统硬件电路具有固化和非灵活的特性，无法动态适应这种灵活的软件识别向硬件识别迁移需求。

为解决上述2个问题，网络管控系统提出软硬件循环迭代迁移处理方法。以柔性可重构技术实现软件功能的硬件迁移，突破了传统的刚性封闭系统的构建方式；以关键词联合匹配构件组串联前置于深度挖掘组件，将带“指纹”特征的数据分流，仅使无“指纹”特征数据进入软硬件循环迭代流程，降低了对识别速率的要求；以软硬件循环迭代结构实现软硬件协同递归式最优通联特征遴选，突破了无“指纹”特征业务的纯软件通联特征遴选和业务识别模式，在现有器件支持下，可将迁移前每轮次兆比特级处理速度提升至迁移后的每轮次吉比特级。

3.1 软硬件循环迭代迁移结构

图2是软硬件循环迭代迁移处理结构，该结构中，关键词联合匹配构件组前置于迁移构件组，负责将具有“指纹”特征的数据分流，约减后的未识别流量（循环迭代迁移流量）输入由迁移构件组、交换组件、数据分发组件和后台软件组成的软硬件循环迭代迁移流程。其中，交换组件实现深度挖掘到组件和数据分发组件的数据交换，为组件间数据交互提供可靠、可控的通路。数据分发组件将交换组件送来的数据无阻塞地分发至各个后台软件子系统。迁移构件组的迁移稳态构件在软硬件循环迭代完成最优通联特征遴选、业务识别模板确认和业务识别验证后，实施硬件电路重构配置，实现“软件→硬件”迁移，后续输入的该业务数据可完全由稳态构件硬件识别，并通过迁移阀将识别结果传送出去，后台软件系统和迁移暂态构件将不再参与该业务识别。迁移暂态构件与特征学习构件联合后台软件具体完成最优通联特征遴选、业务预识别模板自学习和业务识别验证。特征学习构件完成联通特征的线速提取、存储、更新和剔除，并将最新的学习结果反馈给软件。迁移暂态构件包含多个业务预识别模板，每一个模板对应一种业务的“预热”识别方法，具体就是指该业务识别使用了哪些特征，如何综合各种特征判定该业务类型，迁移暂态构件基于遴选特征、预识别模板进行业务“预热”识别

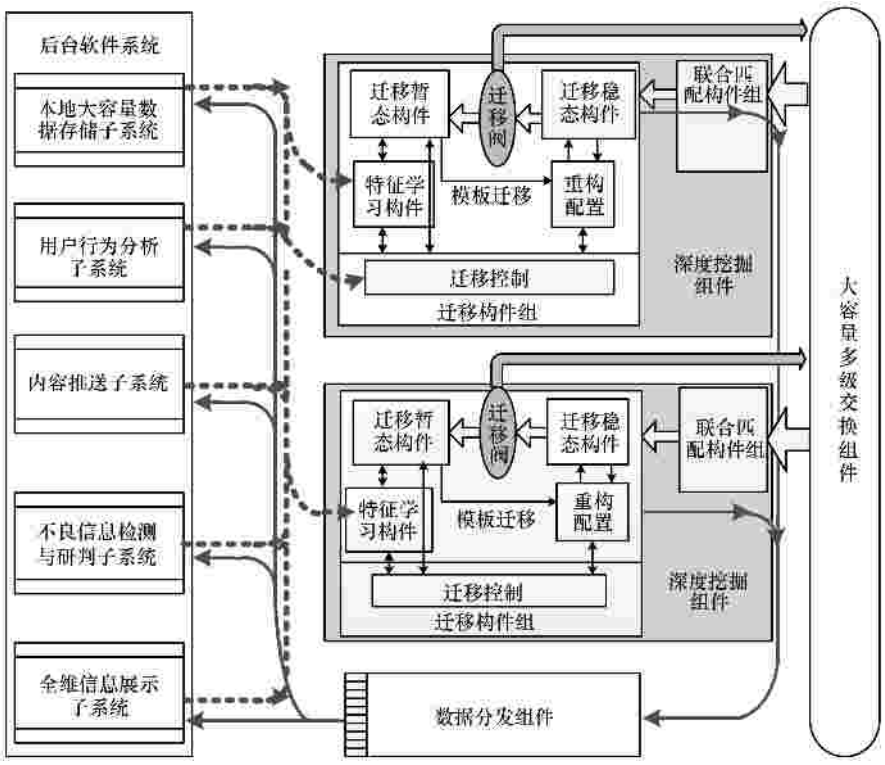

图 2　软硬件循环迭代迁移处理结构

和验证。这里，“预热”识别是指联通特征遴选未完成，业务识别模板未确定，迁移未实施前由迁移暂态构件基于当前特征和预识别模板完成的业务识别。重构配置模块为迁移稳态构件提供识别模板及参数支持，控制电路重构。互联网业务识别与管控系统中，深度挖掘组件由大规模 FPGA（field programmable gate array）构成，在具备高速工作能力的同时提供了硬件重构基础[10]。

3.2　软硬件循环迭代迁移流程

图 3 是软硬件循环迭代迁移处理无“指纹”特征或指纹特征难获取业务的过程。该过程为分组触发处理过程，当新数据分组输入后，首先经过关键词联合匹配构件组进行识别，分流出具有“指纹”特征的业务流，从而降低后续的软硬件循环迭代处理压力。关键词联合匹配构件组未识别的数据分组进入迁移稳态构件进行识别，若识别成功，表明该数据分组属于已经完成“软件→硬件”迁移的某业务类型，识别过程结束，若未识别，则表明该数据分组所对应的业务特征未开始或者未完成业务特征遴选、业务识别验证和迁移，需进入迁移暂态构件进行“预热”识别，由此进入软硬件循环迭代迁移处理。首先由硬件迁移构件组的特征学习构件在原存特征信息的基础上，线速完成当前分组的流识别和通联特征信息的提取及更新。随后，迁移暂态构件基于更新的通联特征统计结果和软件通告的业务预识别模板完成新一轮的业务“预热”识别，并根据所采用的早期识别算法[11~13]判定本轮“预热”识别是否标志整个“预热”识别结束：若未结束，则将本轮更新的通联特征通过交换组件和数据分发组件输出至后台软件，由软件进行业务特征的智能学习、遴选、“疑似”特征降维和预识别模板自学习，并将本轮降维后的疑似“特征”和预识别模板通告给硬件迁移构件组的特征学习构件，由该构件在原来流信息的基础上增建新特征，并线速完成特征信息的维护和更新，从而完成新特征的软件筛选、硬件同步，本轮软硬件循环迭代迁移处理结束，等待新数据输入触发新一轮循环迭代迁移处理；若结束，即本轮“预热”识别是最后一轮，则该结束信息联合流量迁移门限共同判定是否迁移，若迁移，则对迁移“稳态”构件进行重构配置，并将迁移“暂态”构件最后接收到的预识别模板作为最终的识别模板迁移到迁移“稳态”构件中，完成整个“软件→硬件”迁移，此后，该业务的识别完全由硬件实现，软件不再参与，除非该业务又出现新的变化，需要重新启动新一轮迭代和迁移。

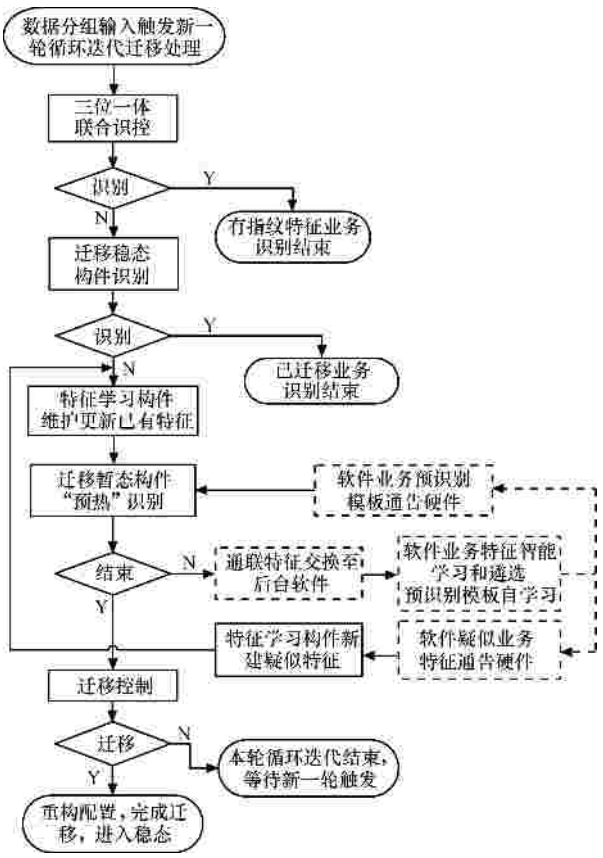

图 3 软硬件循环迭代迁移处理过程

3.3 硬件循环迭代迁移实现

软硬件循环迭代迁移处理基于 FPGA 动态可重构技术实现，其可重构迁移结构如图 4 所示。

迁移结构由迁移暂态构件、迁移稳态构件、迁移控制构件和重构配置构成。其中，迁移暂态构件由若干预匹配构件构成，迁移稳态构件由若干匹配模板构成。迁移暂态构件进行“预热”识别，迁移条件满足后，匹配模板和特征参数通过 FPGA 动态重构配置给稳态构件，将迁移稳态构件中的通用匹配模板配置为针对各类具体业务识别的专用匹配模板构件，完成硬件电路的更新和使能，从而具备了新业务或无“指纹”业务大流量情况下的高速识别能力。

系统将无“指纹”特征匹配功能设计成标准的 FPGA 匹配模板重构件，使这些匹配模板构件具有统一的内部接口、外部存储接口和相同的时序关系。如图 5 所示，每一个匹配模板构件都被划分在 FPGA 的一个区域内，占用该区域的逻辑资源、存储资源和管脚资源等，所占用资源的总量与无“指纹”特征匹配的复杂度相关。需要指出的是，由于各个匹配模板构件都是独立实现的（占用独立的 FPGA 区域，并且各个模块间不存在相互通信），因此在系统动态地增加新的匹配模板构件时，其他业务类型的匹配模板和整个系统的运行将不受影响。

3.4 硬件循环迭代迁移小结

软硬件循环迭代迁移处理具有如下优点。

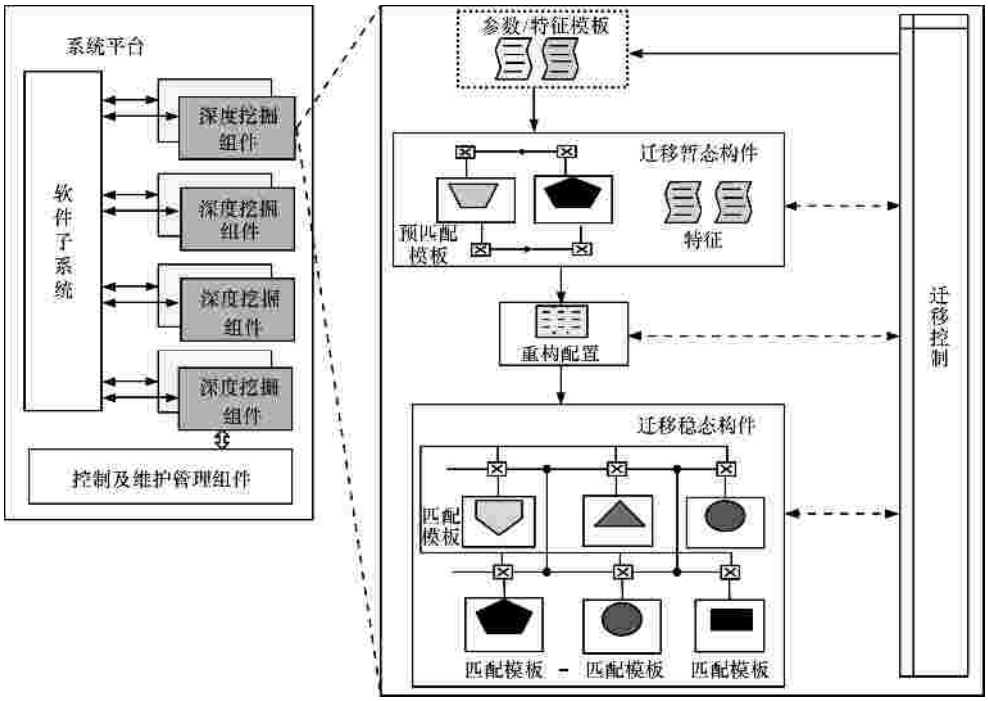

图 4 柔性可重构迁移结构

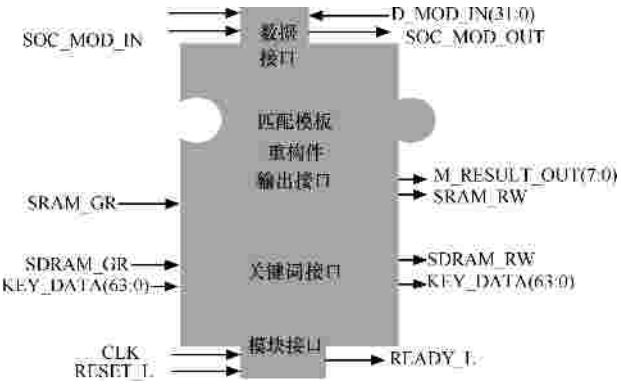

图 5　匹配模板重构件

1) 关键词联合匹配构件组串联前置于迁移构件组，分流了带“指纹”业务的数据，降低循环迭代迁移对识别速率的要求。

2) 未迁移情况下，业务识别由软硬件协同完成处理，硬件的特征学习构件线速完成特征信息的提取、更新，软件基于硬件提供的特征信息完成机器自学习和特征遴选，处理速率可达吉比特级，远超过纯软件的兆比特级处理速度。

3) 迁移结束后，业务识别完全硬件处理，可达到线速处理目标。

4) 在不改变硬件系统结构和开发新硬件平台的情况下，实现了新业务识别的“软件→硬件”动态迁移。

软硬件循环迭代迁移处理面临的技术难点包括以下 3 点。

1) 特征学习构件通联特征维护更新难。为了快速、高效地获得新型业务的固有特征，做到近似“零耽搁”识别，特征学习构件需在线完成业务流量统计特征及用户行为特征的提取，线速建立基于流的特征详细记录，对新输入数据分组线速识别流和更新流特征详细记录，而目前支持高速通联特征维护和更新的存储器件中，速度快的容量不满足要求，容量大的速度又不满足要求[14,15]。

2) 通联特征的降维难。待识别的网络流量数据集由网络中的流组成，每个流由一系列的通联特征来描述，这些特征多达 248 种，如此多的流特征中存在着大量不相关和冗余的特征，为了提高分类性能，需要去除这些冗余特征，实现数据的降维。若降维程度不够，降维后的特征种类仍过多，硬件存储空间的需求将过大，可实现性将降低；而若降维程度过大，降维后的特征种类过少，分类的准确度将降低，业务识别的准确性也将降低。

3) 硬件公平抽样难。当前器件水平下，OC-768 (40Gbit/s)链路速率下的逐分组流识别、流特征存储和更新的实现难度大、代价高。实际上，采用抽样方法对流特征进行统计和更新是目前普遍使用的一种高性价比做法，硬件常用的抽样方法是简单随机抽样和固定概率抽样，例如 Cisco 在 NetFlo 的“1 out of N”的静态抽样策略[16]，可以较好地反映“大象”(elephant)流特征，但会严重影响“老鼠”(mouse)流特征，而考虑了大流和小流之间抽样公平性的各类算法，例如草图指导的抽样 (SGS, sketch guided sampling)算法在高速网络下难以实现[17]。

4　结束语

作为一种新型网络设备，网络管控系统至今未有适合其应用的成熟体系结构模型，沿袭的路由器结构存在明显的体系结构不适应问题。本文提出了网络管控系统的软硬件交换式松耦合协同处理结构，摒弃了传统核心路由器数据平面、控制平面分离结构，实现了基于交换组件的软硬件融和式协同数据处理，适应了业务识别和管控的线速、深层次智能处理需求；提出了软硬件循环迭代迁移结构和方法，突破了以往 DPI 设备的刚性封闭系统构建方式，以柔性可重构技术实现软件功能的硬件迁移，满足了新业务识别的“软件→硬件”迁移需求；以软硬件循环迭代结构实现软硬件协同递归式最优通联特征遴选，突破了无“指纹”特征或指纹特征很难获取业务的纯软件识别模式，将传统上每轮次兆比特级处理速度提升至每轮次吉比特级。

参考文献：

[1] Cache logic[EB/OL]. http://www.cachelogic.com, 2008.

[2] A new generation of high trusted network[EB/OL]. http://www.863.gov.cn/news/0907/05/0907053613.htm, 2009.

[3] YD/T, 深度分组检测设备技术要求[S]. 2009.
YD/T, Technical Requirement of Deep Packet Inspection Device[S]. 2009.

[4] DPI 深度分组检测技术及其作用[EB/OL]. http://www. huawei.com/cn/products/datacomm/catalog.do?id=2146, 2007.
Technology and effect of deep packet inspection[EB/OL]. http://www.huawei.com/ cn/products/datacomm/catalog.do?id=2146, 2007.

[5] 李玉峰,兰巨龙,薛向阳. 面向三网融和的统一安全管控技术[J]. 中兴通讯技术, 2011,17(4):23-28.

LI Y F, LAN J L, XUE X Y. Common security and control framework in tri-network convergence[J]. ZTE Technology Journal, 2011,17(4):23-28.

[6] QI Y, FONG J, JIANG W, *et al*. Multi-dimensional packet calssification on FPGA: 100Gbps and beyond[A]. Proceedings of Internation Conference on Field-Programmable Technology[C]. Beijing, China, 2010. 241-248.

[7] SUNG J S, KANG S M, LEE Y, *et al*. A multi-gigabit rate deep packet inspection algorithm using TCAM[A]. Globecom 2005-IEEE Global Telecommunications Conference[C]. St Louis, Missouri, USA, 2005. 453-457.

[8] GAO M, ZHANG K N, LU J H. Efficient packet matching for gigabit network intrusion detection using TCAMs[A]. Proceedings of the 20th International Conference on Advanced Information Networking and Applications[C]. Vienna, Austria, 2006. 249-254.

[9] MOORE A, ZUEV D, CROGAN M. Discriminators for Use in Flow-Based Classification[R]. RR-05 13 Department of Computer Science, University of London, 2005.

[10] **李玉峰,邱菡,兰巨龙**. **可重构路由器研究的现状与展望**[J]. **中国工程科学**, 2008,10(7):82-89.
LI Y F, QIU H, LAN J L. Status quo and outlook of reco figurable research[J]. Engineering Sciences, 2008,10(7):82-89.

[11] BERNAILLE L, TEIXEIRA R, SALAMATIAN K. Early application identification[A]. The 2nd ADETTI/ISCTE CoNEXT Conference[C]. Lisboa, Portugal, 2006. 456-468.

[12] BERNAILLE L, TEIXEIRA R. Early recognition of encrypted applications[A]. Proc of PAM[C]. Louvain-la-neuve, Belgium, 2007. 165-175.

[13] HUANG N F, JAI G Y, CHAO H C. Early identifying application traffic with application characteristics[A]. Proc of ICC[C]. Beijing, China, 2008. 88-92.

[14] IYER S, KOMPELLA R R, MCKEOWN N. Techniques for fast packet buffers[EB/OL]. http://www. comsoc.org/tcgn/conference/ gbn 2001/iyer-presentation.pdf,2001.

[15] IYER S, KOMPELLA R R, MCKEOWN N. Analysis of a memory architecture for fast packet buffers[A]. Proc of IEEE High Performance Switching and Routing[C]. Dallas, Texas, 2001.368-373.

[16] Cisco netflow[EB/OL]. http://www.cisco.com/warp/public/732/Tech/netflow, 2005.

[17] KUMAR A, XU J J. Sketch guided sampling: using on-line estimates of flow size for adaptive data collection[A]. IEEE Infocom[C]. Barcelona, Catalunya, Spain, 2006. 1-11.

作者简介：

李玉峰（1975-），男，山东烟台人，博士，国家数字交换系统工程技术研究中心讲师，主要研究方向为路由与交换、网络业务识别与控制等。

邱菡（1981-），女，湖北随州人，博士，解放军信息工程大学讲师，主要研究方向为网络信息安全等。

刘勤让（1975-），男，河南周口人，博士，国家数字交换系统工程技术研究中心副教授，主要研究方向为高速宽带信息网络、SoC 技术。

兰巨龙（1962-），男，河北张北人，博士，国家数字交换系统工程技术研究中心教授，主要研究方向为高速宽带信息网络。