基于DPI内容分析的业务流量识别技术

上网行为管理、流量管理与下一代防火墙等设备的技术核心基础在于应用协议识别，离开了高效精确的协议识别，网络应用的阻断、流控与审计等功能都无从谈起。

信达网安 以DPI（Deep Packet Inspect，深度包检测）技术为核心，结合基于报文内容及基于行为特征等核心技术，实现网络中应用的自动识别和智能分类，如图1所示。到目前为止，已经支持几十组与上千多种网络应用的识别。

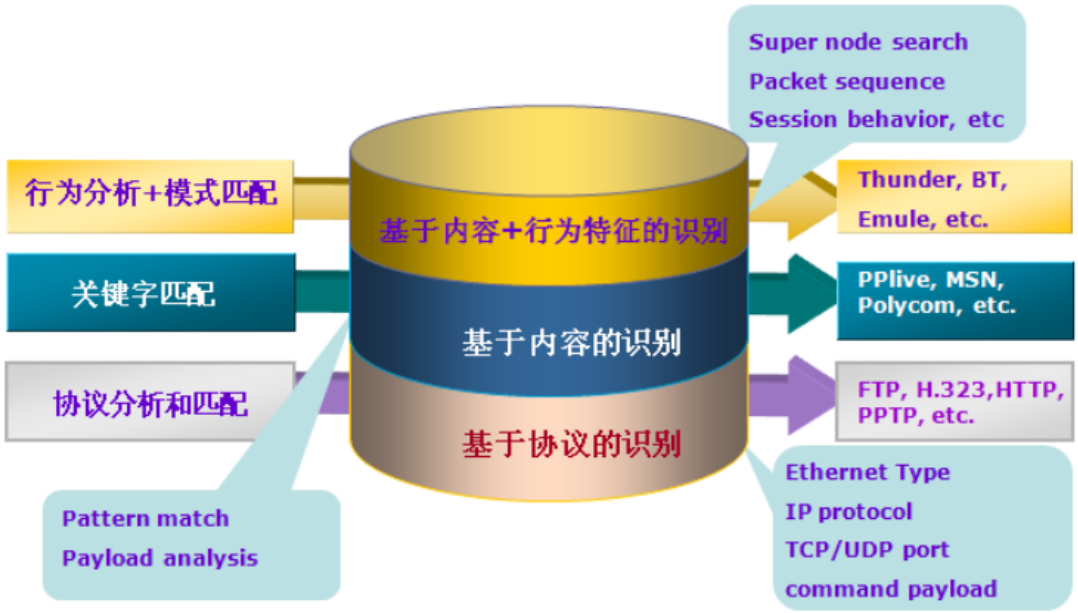


图1 DPI识别技术

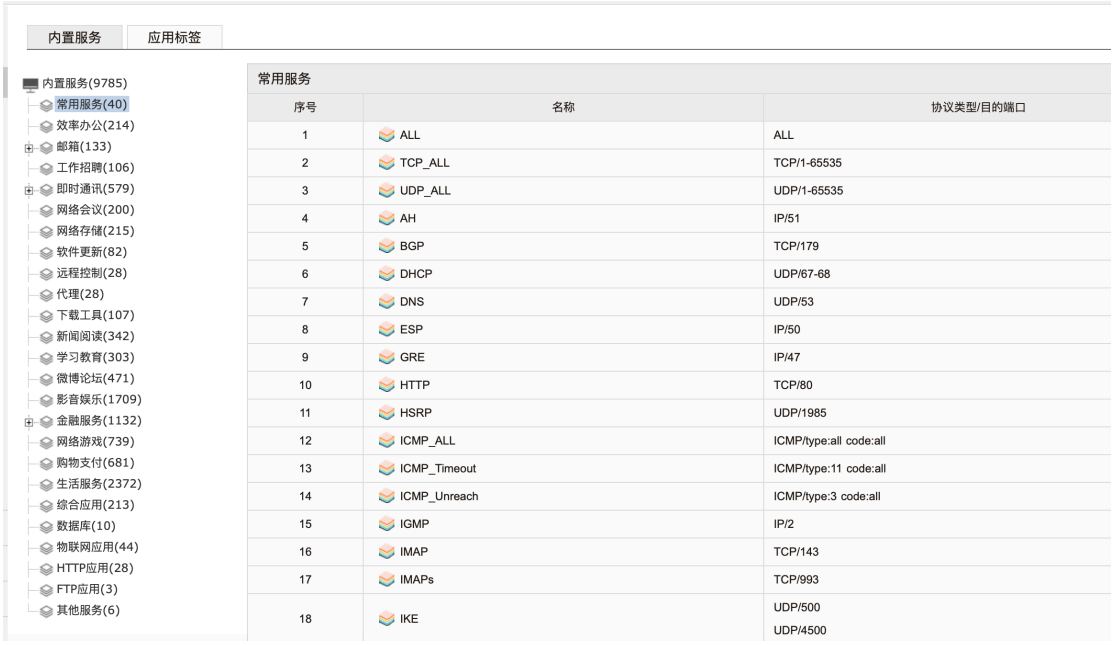
到目前为止，DPI 已经支持 9000 多种协议的识别，如下所示：

- a) 应用规则识别库：该库由应用规则研发团队定期维护，保证库处于最新状态；该库支持 9000 种以上网络主流应用，能识别 125 种以上 IM、100 种以上 P2P/P2P 流媒体、600 种以上游戏、 30 种以上 OA、30 种以上网银、1000 种以上金融行情软件、45 种以上金融交易软 件、13 种木马、30 种以上代理软件和 590 种以上移动 APP，物联网应用 40 种以上，涵盖主流的网络应用；
- b) 自定义规则：支持管理员自行定义新规则；
- c) 智能识别：种类泛滥的 P2P 行为，静态“应用识别规则”已经捉襟见肘，通过

P2P 智能识别技术，识别出不常见、未来可能出现的 P2P 行为，进而封堵、流控和审计。

✧ 通过强大的应用识别技术，无论网页访问行为、文件传输行为、邮件行为、应用行为等，都能帮助组织实现对上网行为的封堵、流控、审计等管理。

按应用类别分类如下：



基于内容分析的业务流量控制区别于传统L3/L4协议IP报文过滤与控制，为了提高识别这类流量分析过滤带宽，信达网安 采用内核匹配数据面与管理面相分离，用户控件通过接口加载特征库，并输出相应的网络应用/应用分组，以供二次开发引用，而内核高效匹配来完成协议识别。如图2所示，这种用户空间应用只分析加载特征而又内核高效匹配方法，使得信达网安 网关产品业务识别准确，流量无损，产品在业界颇具竞争力。

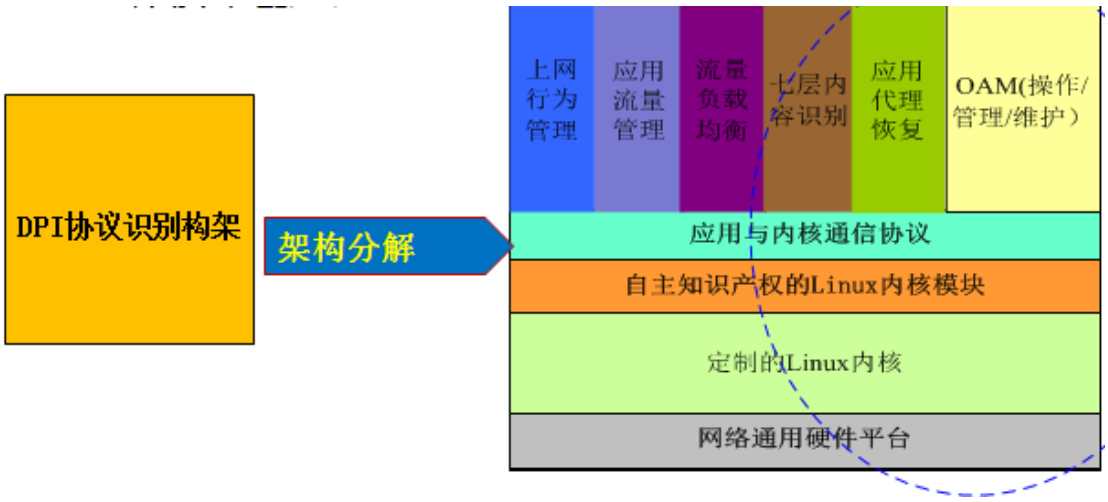


图2 DPI协议模块构架

以识别浏览器访问类的网络应用为例，如图3所示，说明如下：

首先，其处理流程大致如下：针对数目众多、协议规则不同的各种网络应用，信达网安 在特征库加载时，将冗长特征库链表根据五要素多次折叠成匹配矩阵，也就是在内核表项中为每种软件/协议依据图3描述的五要素，即报文方向(REQUEST/RESPONSE)、协议类型TCP/UDP、四层端口(1-65534)、HTTP协议种类(POST/HEAD/PUT/GET)与HTTP协议内容(HOST/USERAGENT/REFFER/URL/CONTENT)建立不同的匹配列表。这样通过这些五要素来分类，特征库匹配列表的深度大幅度降低，同时对关键字的内容采用了著名的Wu-manber多模匹配算法，严格地讲，匹配过程中我们抛弃了正则表达式，而采用了Wu-manber字符块匹配算法，不存在深度值较大的匹配链表，极大地提高了匹配效率。

其次，我们DPI是以内核conntrack为单位来进行协议识别的，不另行建立相关的流表，仅仅在conntrack加入一些控制字段，这在二次开发的说明手册中已经有详细的描述，减少了一次不必要的hash表的维护操作，也是出于匹配效率的考虑。

最后，建立协议识别的节点跟踪cache及其时效性，针对某些UDP应用协议的特点，也极大地提高了应用协议的二次识别的效率。比如，通过深度DPI识别出的协议，在一定的时间内它的目的IP/端口保持不变，对该时效范围内同一IP/端口，立即可以识别成同一UDP应用协议。这也极大地提高了某一类协议的识别效率。

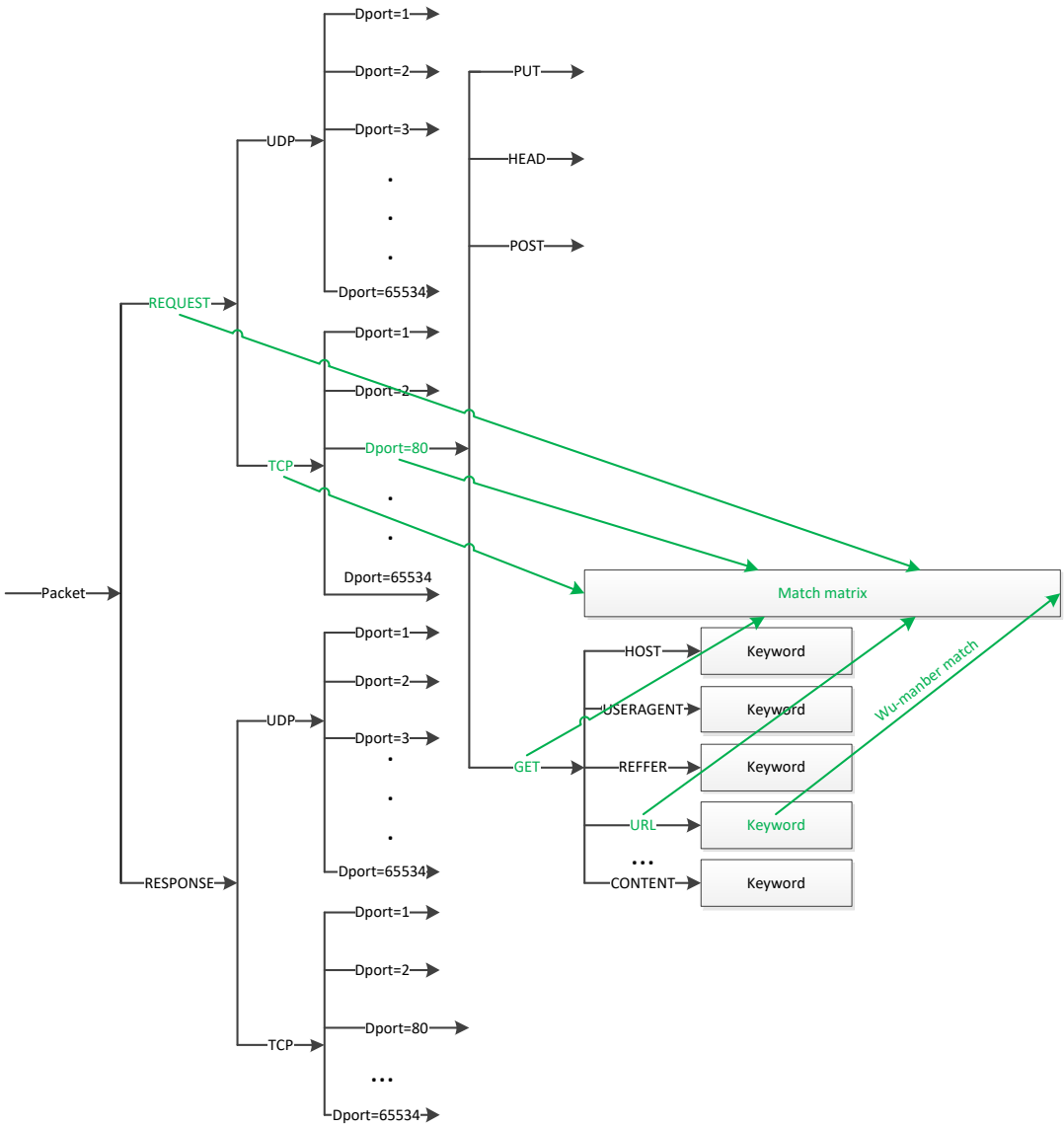

Dport=1
Dport=2
Dport=3
Dport=65534
UDP
PUT
HEAD
POST
REQUEST
Dport=1
Dport=2
Dport=80
TCP
Dport=65534
Packet
Match matrix
Dport=1
Dport=2
Dport=3
Dport=65534
UDP
HOST
USERAGENT
REFFER
URL
CONTENT
GET
Keyword
Keyword
Keyword
Keyword
Keyword
Wu-manber match
RESPONSE
Dport=1
Dport=2
Dport=80
TCP
Dport=65534

图3 高效的匹配矩阵