网络神探动态域名搜索备案-关键技术详细设计

1 简介

动态域名搜索服务通过预先设置的 IP 地址库进行搜索，对 IP 地址段内每一 IP 逐一解析，获取 IP 对应域名信息后入库存放，通过智能解析服务与系统备案基本资料库进行数据比对获取相应备案关联信息，并进行校验以及捕获域名首页 html 等相关资料。

2 问题描述

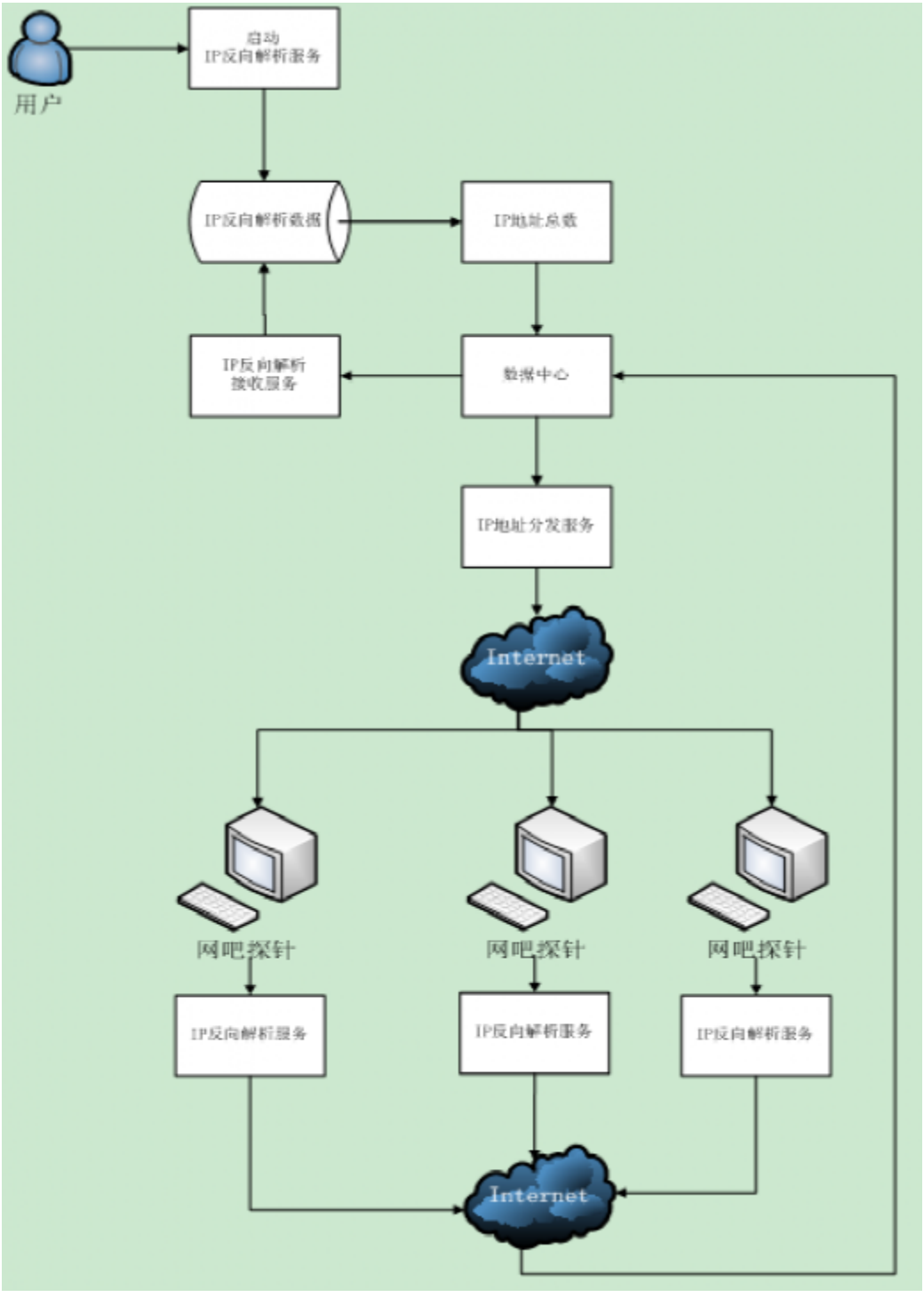
IP 反向解析服务通过第三方网站进行解析获取数据，存在频繁访问网站后 IP 地址被屏蔽的情况；这情况导致服务无法正常运行与解析速度非常缓慢问题；

目前杭州运行的 IP 地址总数 4491152 条，（设计要求 700 万 IP 地址），每个 IP 解析平均需要 20 秒（均值），6 台服务器，4 个进程进行解析，450 万需要 43 天；700 万需要 67 天;

3 解决方案

针对这种情况, 可用网吧终端资源，建立分布式查询技术，将原来单节点服务器的查询任务分解到各个网吧终端探针，即对杭州市 IP 地址段拆解到各个网吧进行反向解析获取数据，能够显著提升解析速度获取解析结果，并减少被 IP 地址反向查询网站屏蔽 IP 地址的概率；

基本逻辑架构如下：



本改进涉及如下关键技术：

1）IP 地址段解析任务分发

2）探针端 IP 地址反向解析

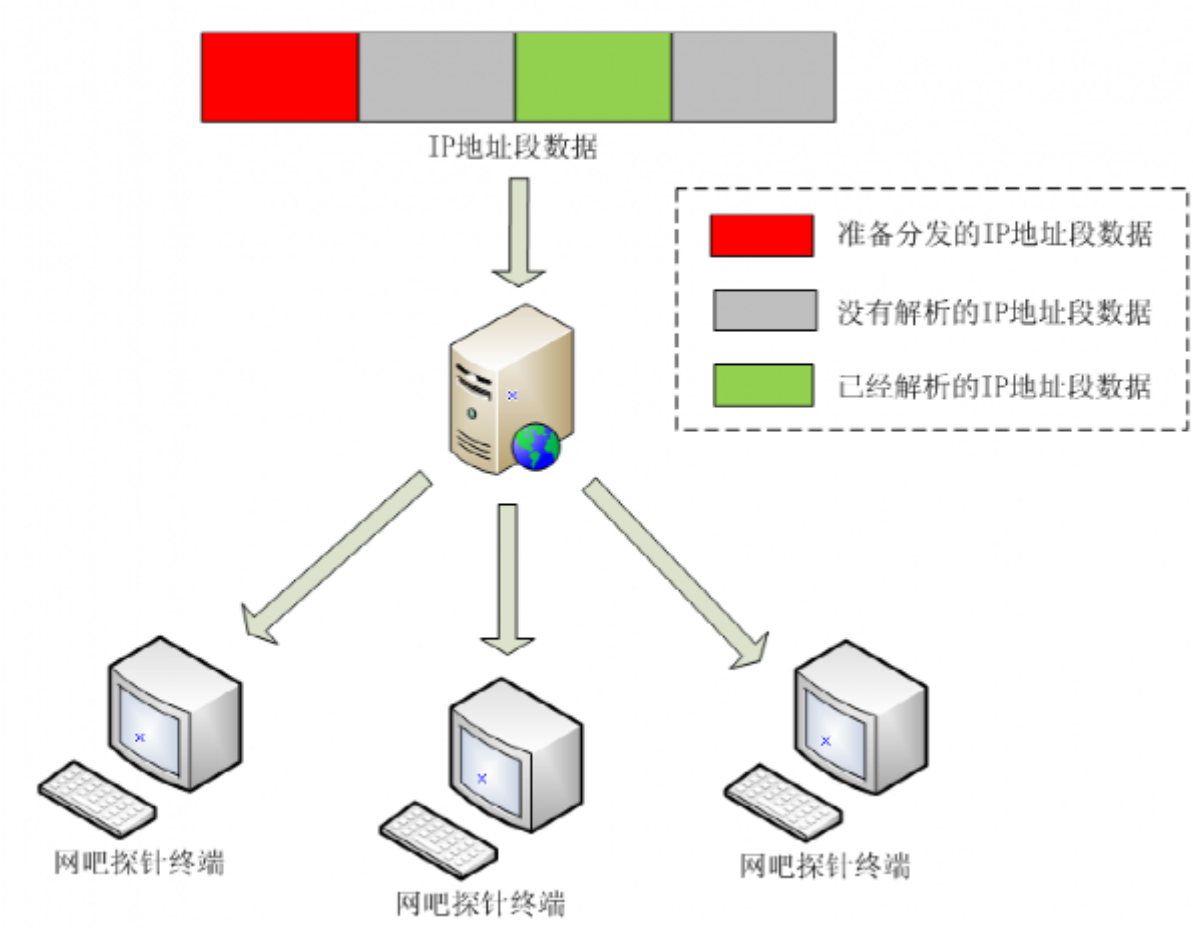
3）解析结果回传

3.1 IP 地址段分发

IP 地址段分发采用客户端“PULL”技术，即在探针端部署一个客户端，由该客户端主动从数据中心服务器获取还没有执行反向解析的 IP 地址段，并将其下载到客户端执行解析，并把最终结果返回给服务器。具体分发模型见下节。

3.1.1 IP 地址分发模型

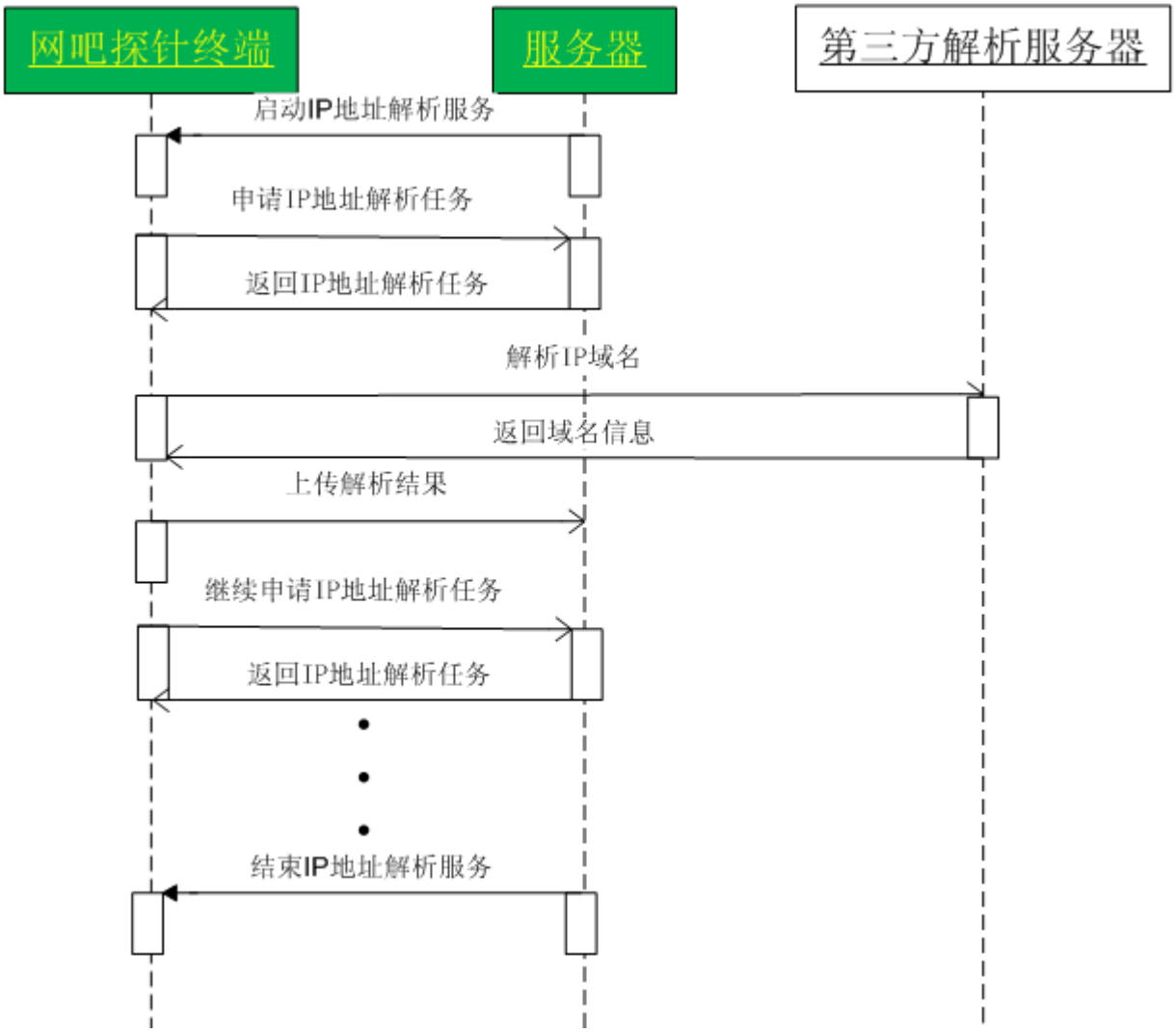
由于客户端的不确定性，不一致性以及不稳定性，为了将所有 IP 地址分解到网吧探针客户端去执行，必须要有一个高效，可靠的模型来保证所有客户端在执行速度和稳定性方面都能达到最佳的状态。下图是新方案的分发模型设计：



如上图所示，分发服务器将所有 IP 地址段数据均分为大小相等的数据块，每次探针终端请求 IP 地址解析时，服务器从没有解析的地址数据块中分配一个给终端进行解析，当该探针分析完并将结果回传到服务器时，服务器将其状态标识为解释完成，如上图绿色所示。

3.1.2 IP 地址分发流程

依据上节描述的分发模型，如下序列图详细地描述从服务器端启动一次全 IP 地址段解析的整个过程:



如上图所示，当服务器发指令给每个探针终端启动解析任务后，所有探针终端主动向服务器申请地址解析任务，并将结果返回给服务器，然后继续申请下一个任务，直到所有地址解析完为止，然后服务器发指令给客户端，让其结束查询服务。

3.2 探针端 IP 地址反向解析

目前反向解析是利第三方免费网站进行，以下是依据 IP 地址解析数据返回情况做了一次统计排名；依据返回数据比较采前 5 个网站做解析　务采集数据源；

- http://tool.chinaz.com/Same/?s
- http://dns.aizhan.com
- http://ip.valu.cn
- Myip.cn
- http://www.114best.com/ip/114.aspx
- http://www.dirs.cn/

- http://whosonmyserver.com
- http://ip.aadata.net/index.asp
- http://www.293.net/

（以下不可用）

- Yeugetsignal.com（访问不到网站）
- Zhanmama.com（返回结果很少）
- Webhosting.info（需要注册用户）
- Sundns.com（需要验证码）

具体解析步骤见以下几节。

3.2.1 解析过程

1）请求反向解析网站，获取反向解析网站返回内容

通过反向解析网站 URL 规则，传入要解析的 IP 地址，通过服务直接请求获取反向解析网站 Response 的内容；

2）提取反向解析网站返回信息中 IP 和域名对应信息

反向解析网站 Response 返回的内容以字符串形式进行正则表达式规则匹配，提取符合规则的 IP 与域名对应信息；

3.3 解析结果回传

3.3.1 客户端

3.3.2 服务器端

服务器端接收到客户端的解析结果后，将其保存到反向解析数据库（见下节），入库时以 IP 和域名作为联合主键，判断记录是否存在，存在则更新最后时间，不存在则插入新记录。

反向解析数据库

表名：tIPAddress(IP 地址段表)

字段名称	字段类型	说明
iIPAID	PK int not null	自增 ID
lIPResID	Bigint null	IP 地址关联单位 ID
sIPBegin	Varchar(50) not null	开始 IP 地址
sIPEnd	Varchar(50) not null	结束 IP 地址
IsDispatch	Smallint not null	是否分配
iSearchCount	Int not null	解析数量
iIPBegin	Bigint null	开始 IP 地址
iIPEnd	Bigint null	结束 IP 地址

表名：tIPDomains（IP 域名记录表）

字段名称	字段类型	说明
lID	PK bigint not null	自增 ID
sIP	Varchar(50) not nul	IP 地址
iIP	Bigint null	IP 地址（数字格式）
sDomain	Varchar（500）not null	域名地址
dUpdateTime	Datetime not null	更新时间
iStatus	Int not null	验证信息(0:未验证； 1：验证通过;2：验证不通过)
iType	Int null	域名类型 1:一级域名; 2：二级域名
iIsNew	Int not null	是否新增记录（1：是）
sTimeStamp	Varchar(50) null	时间戳

表名：tIPNoDomain（未获取域名的 IP 记录表）

字段名称	字段类型	说明
iID	PK bigint not null	自增 ID
sIP	Varchar(50) not nul	IP 地址
iIP	Bigint null	IP 地址（数字格式）
dAddTime	Datetime null	添加时间
sTimeStamp	Varchar(50) null	时间戳