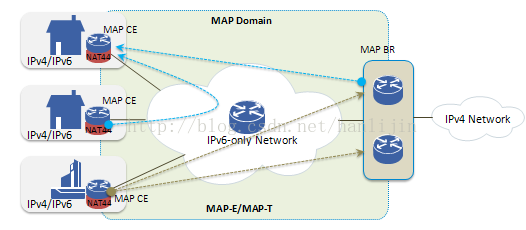
**城域网IPv6过渡技术——MAP技术**

在IPv4向IPv6网络演进的浪潮中，存在着两对主要矛盾的较量，一对是IPv4地址短缺和IPv4业务蓬勃发展之间的矛盾，另一对是IPv6海量的地址空间和IPv6应用的匮乏之间的矛盾。在IPv4方面通过地址复用（A+P)方式似乎缓解了IPv4快速消耗的压力，但是NAT设备投入巨大，各类业务应用也或多或少受到影响。在IPv6发展方面，用户、ICP、ISP以及运营商对IPv4地址枯竭的敏感度不一致，从而导致IPv6产业链发展不平衡，各方面在积极推动IPv6发展的同时都或多或少存在顾虑。同时两对矛盾又互相制约，IPv4地址共享机制似乎又减缓了IPv6产业链的发展，IPv6产业链的不断发展似乎又在考验着IPv4地址共享机制的部署规模。

为了保持IPv4业务持续性和促进IPv6产业的发展性，**4over6**场景凭借其兼顾IPv4业务和IPv6发展的特点，成为长期演进方案研究的焦点。在4over6场景中，**MAP**技术结合了**无状态**和**双重翻译**/**封装技术**，成为目前**IETF**关注度最高的解决方案。**MAP**（Mapping Address and Port）技术是指**无状态地对地址和端口进行复用**，根据报文格式又分为**双重封装MAP-E**和**双重翻译MAP-T**两种。**MAP**技术文稿目前是**IETF**的**WG draft**状态（截止2013年11月）。

MAP技术定义了在IPv6-only的网络中承载IPv4和IPv6业务无状态地址封装/翻译的机制。**MAP-CE**和**MAP-BR**作为**边界设备**划定了MAP Domain的区域，IPv4业务流仅存在于**MAP Domain**之外。



我个人认为，MAP技术的本质特征可以简单总结为四个关键字，即**共享、无状态、分布、优化**。

* **共享**：采用A+P理念共享Public IPv4地址，即通过TCP/UDP层的端口号资源来扩展Public IPv4资源，分配Port-Set受限的IPv4地址**实现多个用户共享**同一个IPv4地址。
* **无状态**：无状态方式实现IPv4 A+P到IPv6地址映射。即通过MAP Rule实现IPv4地址+端口与**IPv6地址的无状态映射**，解决CGN维护地址映射状态的问题。
* **分布**：NAT44分布到**MAP CE**实现。即秉承NAT卸载的理念，利用MAP CE的资源实现**分布式**NAT，解决集中式NAT44性能瓶颈的问题。
* **优化**：优化流量转发和提高网络可靠性。即通过配置**MAP Rule**实现MAP CE间流量直接转发，通过**MAP-BR Pool**实现**负载分担**，同时由于无状态映射的特点，MAP BR故障切换时无session同步需求，网络可靠性更高。

MAP技术的实现需要考虑以下几个关键问题：

1. IPv4地址和端口与IPv6地址之间遵循什么**映射规则**？
2. MAP域中是如何来**规划映射规则**？又是如何进行**规则下发**的？
3. MAP-BR的Pool是如何实现？可靠性如何提升？
4. MAP场景下业务流量的报文格式是什么样的？

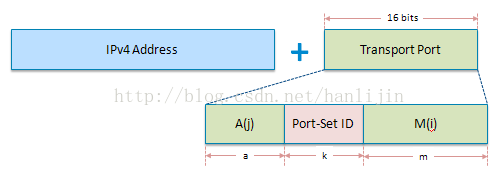
接下来我将逐一进行阐述。

**IPv4地址和端口与IPv6地址之间遵循什么映射规则**？

这是MAP技术要解决的第一个问题，也是MAP技术的基础。

首先，从IPv4地址和传输层port来看，32bit的IPv4地址容量有限，16bit的传输层端口（Transport port)目前使用不多，因此借用传输层端口来扩展IPv4地址是一个不错的方法，这就是“A+P"的概念。

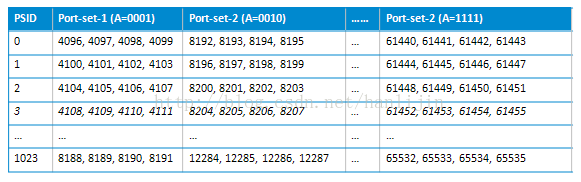
为了确保用户使用的**IP+Port-set**的唯一性，需要将port-set进行有序划分。由于传输层端口中有一些已经**分配给特定的应用使用**（比如80为http，20/21为ftp等)，因此建议分配port-set是需要避免使用此类端口。在MAP中，我们将传输层端口号的16bit划分为三部分，**A、PSID和M**。



* **关于A和a**：因为0-1023是well-known端口区间，在MAP中建议扩展到0-4095（2^12），即建议a(=PSID offset)默认值是4（16-12）。A的取值为**非零值**，M的取值为**任意值**。若**a=0**，则标识全部**port**区间均可以分配。
* **关于PSID和k**：PSID的长度**k**决定了**共享比率**，共享比率值=**2^k**，即传输层端口可以分成2^k份，每份共享给一个**CPE**使用，每个共享此IPv4地址的**CPE**获得一个唯一的**PSID**，即获得一段唯一的**port-set**。
* **关于m**：M域的长度m决定了**port-set**中**端口**的连续长度，连续长度值为=**2^m**。

通过这样的划分，可以获得2^k个port-set，每个port-set由唯一的PSID值标识，每PSID值对应的port-set含有**((2^a)-1)\*(2^m)**的端口号。接下来，通过实例来体验下port划分过程。

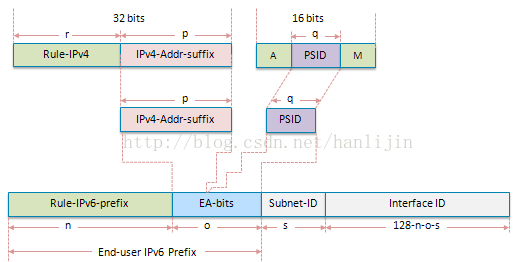
1. 给定共享比例**R**（=2^k=1024）和a(=PSID offset=4);
2. 推算出: a=4,k=10,**m=2;bit**
3. 列出PSID值和对应的**port-set**，



因此，对于给定的唯一PSID值，即获得一组port-set。

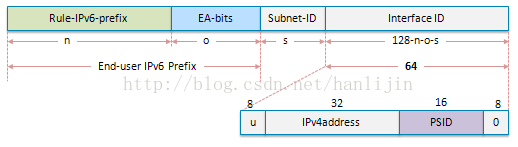
然后，再考虑如何将**IPv4**地址和**port-set**与**IPv6地址**建立强关联关系，形成无状态映射？

MAP技术是通过IPv4地址的的**一部分**和**端口**的**一部分**特征值嵌入到IPv6地址中实现，IPv4地址选的的特征部分是**IPv4-Addr-suffix**，port-set选择的是PSID，这样将IPv4的地址信息和端口信息与IPv6地址建立了强关联，此乃MAP设计的巧妙之处。



通过上图对IPv4+Port与IPv6地址的映射关系可以看出，不管是MAP-CE还是MAP-BR，只要获得**End-user** **IPv6-prefix**、**Rule-IPv6-prefix**、**EA-bits**、**Rule-IPv4**、**PSID** **offset**几个值就推导出共享的IPv4地址和端口序列。

对于IPv6用户而言，MAP-CE配置了**End-user IPv6-prefix**，可以通过Interface ID来生成用户的IPv6地址，而对于IPv4用户而言，不会生成类似于IPv6地址的链路本地地址，那如何来生成标识IPv4属性的Interface ID，进而合成代表IPv4地址的IPv6源地址呢？MAP技术采用**IPv4 address**和**PSID**的值组合形成**Interface ID**来标识IPv4共享型用户，再与End-user IPv6-prefix一起合成IPv6地址，作为IPv4共享型用户在**MAP Domain**中的唯一标识。



* 对于**IPv4address**字段，若分配了一个共享IPv4地址，则IPv4address字段填写分配的IPv4地址，长度是32比特，如果分配了**IP-prefix**，即分配一段地址给IPv4用户（有可能是个企业用户），则IPv4address字段需要右填0补齐，比如给用户分配**IPv4-prefix**="192.0.2.0/29",（29表示有从左向右29个1=>推算出子网掩码为255.255.255.248）则此处IPv4address字段需要填写"0xC0000200"(十六进制）。
* 对于**PSID**字段，若EA-bits位提取的PSID值**不足16bit**时，**左填0补充**，如:PSID=“**0xAC**”，则**此处**的PSID字段填“**0x00AC**”，如果分配了IPv4-prefix或者独享的IPv4-address时，则没有PSID值可以提取，则此处的PSID字段填写“0x0000”。

至此，MAP域中的IPv4地址共享型用户的IPv4地址和端口已经和IPv6地址建立了**无状态映射所需的强关联关系**，并可以构建IPv6地址在MAP域中唯一地标识此用户。

接下来考虑，要考虑第二个关键问题：**MAP域中是如何来规划映射规则？又是如何进行规则下发的？**

上一节讨论了MAP技术中IPv4 A+P与IPv6地址之间映射规则，这一节接着来回答第二个问题“MAP域中是如何来规划映射规则？又是如何进行规则下发的？”

在MAP技术中有三种**MAP Rule**，**BMR**(Basic Mapping Rule)、**FMR**(Forwarding Mapping Rule)和**DMR**(Default Mapping Rule)，其中DMR在MAP-E中成为Destinations outside the MAP domain。

**第一条是BMR，是必选项**，用于计算MAP CE的IPv4地址和port-set以及IPv6地址。BMR是需要配置在MAP域的每一个路由器上，配置在**MAP-CE**上用于将IPv4用户数据进行NAT44和IPv6翻译/封装，配置在**MAP-BR**上用于将IPv4地址从IPv6报文中解封装/解隧道，以及**将回程流量**的IPv4地址+port进行IPv6翻译和封装后，在MAP域中按照IPv6路由转发到**MAP-CE**上。

BMR的配置的基本参数包括：**Rule-IPv6-prefix**、**Rule-IPv4-prefix**、**EA-bits-length**、**PSID-offset**，这些参数配置在**MAP-CE**上可以计算出共享的IPv4地址和端口序列，以及**MAP-CE的IPv6地址**，举个实例分析下（参考上一节的映射图）。

MAP-CE获得属于IPv6前缀信息：（DHCPv6正常分配）

* End-user IPv6-prefix: 2001:db8:0012:3400::/56

通过BMR下发的信息如下：

* Rule-IPv6-prefix        : 2001:db8:0000::/40 🡺 n = 40
* Rule-IPv4-prefix        : 192.0.2.0/24 (0xC0000200/24) 🡺 r = 24
* EA-bits-length           : 16
* PSID-offset(a)           : 4 (default)

则推导出如下信息（MAP-CE的共享A+P和IPv6地址）

* EA-bit                         : 0x1234 🡸 p+q = 0x1234 共16bit
* IPv4-suffix                 : 0x12 (p=32-24=8) p = 32 – r = 32-24 = 8

现在已知p 8bit 、n 40 bit、o = p+n = 16bit、q 8bit、

根据End-user IPv6-prefix: 2001:db8:0012:3400::/56

56 = n+o=n+(p+q) 🡺 2001:db8:00 为n 一共 40bit ，0x12为p共8bit

PSID                          : 0x34 (q=o-p=16-8=8) 🡺 p+q = 16 🡺 q为0x34 8bit

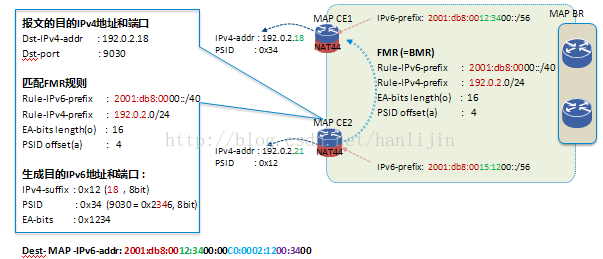
* IPv4-address             : 192.0.2.18 (0xC0000212) 🡺 Rule-IPv4-prefix   + IPv4-suffix = r+p = 192.0.2.0/24 + 0x12 = 192.0.2.18
* **Port-set**                     :  4928-4943, 9024-9039, ......, 62272-62287(a=4, A>0; PSID=0x34)具体计算步骤在笔记
* IPv6-address            : 2001:db8:0012:3400:00C0:0002:1200:3400

End-user IPv6-prefix （7\*16bit->2001:db8:0012:34） + u（8bit->0）+ **IPv4-address （32bit->**0xC0000212**）+ PSID**(16bit->0x3400) + (8bit->0)

在MAP  Domain中可以按照IPv4子网逻辑划分多个sub-domain，每个IPv4子网段作为一个sub-domain，这样在sub-domain中所有MAP-CE配置的MR可简化为一条。每个MAP-CE配置不同的End-user IPv6-prefix和相同的BMR即可。

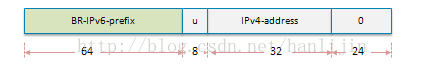
**第二条是FMR，是可选项**，用于在mesh网络中实现MAP-CE间的互访流量直接通信，不需要通过**MAP-BR**中转。这就是说在MAP-CE访问其他MAP-CE时，目的IPv6地址应该是对端MAP-CE的地址。报文的**源IPv6**地址是**BMR**生成的，则**目的MAP-CE**的IPv6地址就成了**FMR**的转换目标。前面也提到**每个MAP-CE的BMR**是相同的（在一定范围），因此FMR和BMR也是可以相同的，即可以将BMR通过配置作为FMR。

举例如下，在MAP-Domain域中配置BMR=FMR，当MAP-CE2用户的IPv4业务访问MAP-CE1用户的IPv4业务时，在MAP-CE2上通过NAT44和BMR生成源IPv6地址，同时通过FMR生成目的MAP-CE1的IPv6地址。



前面两个规则对于MAP-E和MAP-T均适用，在第三条规则上，MAP-E和MAP-T是不同的。

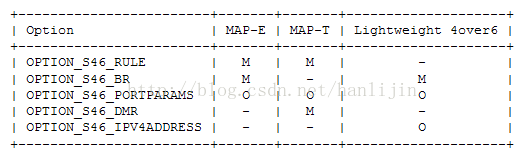
在MAP-T中定了“DMR"，代表报文的目的IPv4地址为MAP Domain域外的地址，通过BR进行转发到域外。DMR含有两个参数**Rule-IPv6-prefix**和**Rule-IPv4-prefix**。Rule-IPv6-prefix的值是MAP BR的IPv6前缀，Rule-IPv4-prefix值是0.0.0.0/0，即在MAP-CE上匹配IPv4路由时作为默认路由适用的映射规则。在使用这条规则后，目的IPv6地址将通过DMR配置的Rule-IPv6-prefix+目的IPv4地址组成目的IPv6地址。



在MAP-E中，第三条规则为Destinations outside the MAP domain，更加明确的表征了规则的意义。由于MAP-E中是IPv6报头嵌套IPv4地址，因此对于报文中目的IPv4地址是MAP域外的地址，只需要再封装一层**BR**的IPv6地址即可，**到达BR**是直接剥掉IPv6报头就可以还原目的IPv4地址。因此在MAP-E中Destinations outside the MAP domain就是配置MAP-BR的IPv6地址。

前面谈了这些规则和属性，那这些是通过**何种协议**进行配置呢？最容易想到的就是DHCPv6属性，但是目前的DHCPv6属性除了配置MAP域所需的End-user-IPv6-prefix之外，其他属性还没有定义，因此就有了draft-ietf-softwire-map-dhcp文档，就是针对MAP-E、MAP-T以及lightweight 4over6（三种协议的特点就是均采用A+P理念）所需的属性对DHCPv6进行了扩展。

在draft-ietf-softwire-map-dhcp中定义了五个DHCPv6 Option，它们分别是OPTION\_S46\_RULE，OPTION\_S46\_BR，OPTION\_S46\_PORTPARAMS，OPTION\_S46\_DMR，OPTION\_S46\_IPV4ADDRESS。另外定了针对MAP-E、MAP-T和LW46三个协议的三个Container，包含的DHCPv6 Option的关系如下。

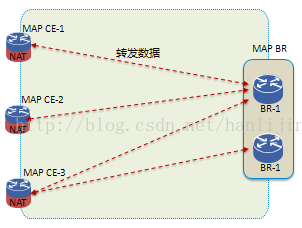


对于每个具体的Option的内容，我就不一一进行介绍了，draft-ietf-softwire-map-dhcp都有明确定义，并且在成为RFC之前都是有可能被讨论和更新的。

本节接着回答MAP技术的第三个问题：“MAP-BR的Pool是如何实现？可靠性如何提升？”

在MAP域中通过将多个MAP-BR放在同一个Pool内实现负载分担和保护倒换的。同一个Pool中的每个MAP-BR配置相同的Anycast IPv6 address和不同的IPv6 address。MAP-CE配置的BR地址是选用的Anycast IPv6 address。

Anycast IPv6 address是IPv6地址族里面一个特殊的地址类型，它是**一组接口**的标识符，送往一个任播播地址的包被传送至该地址标识的接口之一（“路由最近”）。Anycast IPv6 address的这种特点可以实现负载分担或者冗余备份机制。Anycast地址只能作为目的地址，不能用作**源地址**，因此需要为BR另外配置一个IPv6地址。



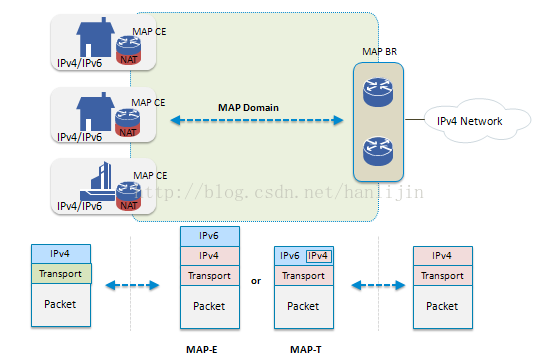
因为MAP属于**无状态**技术，因此对于流量上下行路径不对称的情况仍能正常转发，例如上行CE3到BR-1，下行为BR2-CE3，由于BR-1和BR-2的**MAP Rule**相同，因此数据正常映射和转发；MAP-BR无映射session维护，因此也无热备技术需求。通过Anycast方式配置BR地址，CE通过路由选择最近的**BR**进行转发，实现BR流量的负载分担和提高BR可靠性。

本节回答最后一个问题：“MAP场景下业务流量的报文格式是什么样的？”

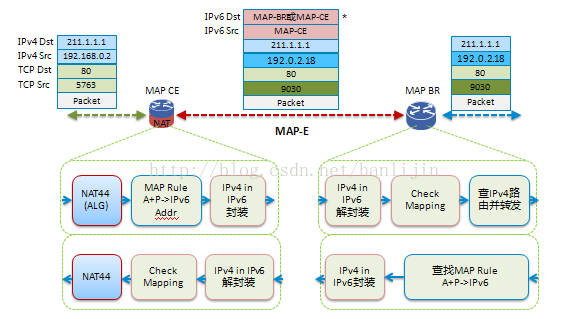
在MAP域中，网络部署IPv6单栈协议。对于IPv6终端的业务流量是采用Native IPv6进行承载。对于IPv4终端的业务流量，需要在MAP-CE和MAP-BR之间(或者MAP-CE和MAP-CE之间）建立IPv6通道，根据对IPv4报文的**封装**方式不同，可以分为**MAP-E**和**MAP-T**两种方式。

MAP-E：封装方式，即将IPv4报文再封装一层IPv6报头，外层为IPv6报头，内层是IPv4报头。

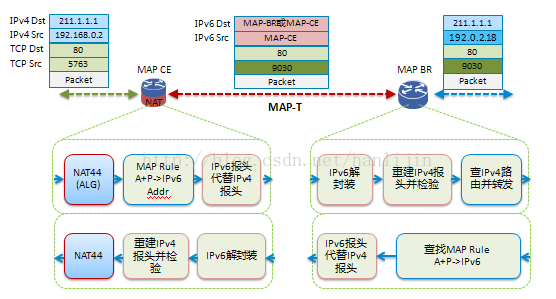
MAP-T: 翻译方式，即将IPv4报头翻译成IPv6报头，只有一层IPv6报头。



通过个人理解，将MAP-E（**双重封装**）的数据处理流程做了一个简单的示意，如下所示。



同理，个人理解，**MAP-T（双重翻译）**的数据处理流程如下图所示。



以上业务处理流程仅为个人理解所设计，欢迎大家跟我讨论。

谈到这里，肯定有人会问：MAP-E和MAP-T哪个更好呢？

其实每个技术都有其特点，对于MAP-E而言增加了一层IPv6报头，增加了报文长度，有可能会引起MTU问题，但是这种封装方式对IPv4的信息是完全保存下来，不破坏IPv4报文的其他属性，即不存在IPv4报文信息的丢失。对于MAP-T而言通过IPv6报头代替IPv4报头，单层报头的报文承载效率高，但是会造成IPv4报头信息的丢失（感觉关键信息已经映射了，只有部分信息有所丢失而已）。