# ACS功能模块定义

## ACS总体方案

根据系统架构对齐将ACS方案设计如图1所示框图。将ACS针对function pooling端口和非function pooling两种模式进行处理。

在非function pooling端口中，不论real mode模式或是isolation mode模式的端口及其下挂的设备全部按照原始的拓扑结构进行呈现且主机域唯一，所以其ACS配置可以按照其主机的配置信息进行访问，仅需处理本芯片内部的虚拟层的ACS配置，端口下挂的设备无需处理。具体详见第二章real mode和isolation mode方案。



**图1 ACS映射方案框图**

在function pooling下无法做到下挂端口共享主机域，所以将所有下挂的Switch设备ACS配置为重定向到上行端口，全部路由到本芯片内部，芯片内部按照虚拟的两级结构对应的各自主机域ACS配置进行逐级匹配路由分发。具体详见第三章function pooling端口方案。

由此可将总流程合并为图2所示。



**图2 ACS路由框图**

DSP ACS配置包括~~V,~~B,R,C,U,E,T及egress\_ctrl\_vector(32bit)总计38bit。

USP 路由配置包含Secondary bus (8bit)，Subordinate bus (8bit)，Non-prefetchable base+limit address(24bit)，Prefetchable base+limit address(88bit)，IO base+limit address(40bit)总计168bit。



## real mode和isolation mode方案

根据系统流程的确定，将ACS独立处理，根据以下表3的分类处理最终ACS将配置将根据配置空间的各种配置信息确定路由的方向。DSP处理如下：



**图3 ACS路由分发模块框图**

本模块并不修改ACS内容的数据和路由的原始数据，添加2bit伴随信息以确定ACS的判断结果给路由模块进行下一步处理。

**表1 ACS路由伴随信息**

|  |  |
| --- | --- |
| 字段定义 | 路由方向 |
| 2'b00 | 默认值 |
| 2'b01 | UP |
| 2'b10 | 违例CPL |
| 2’b11 | P2P |

### 1、ACS Source Validation来源验证

该部分在PLDA 提供的第三方IP DN controller中进行处理，当ACS Source Validation Control Bit被使能时，接收到的Upstream Request的Requester ID中的bus number不在Switch Downstream Port的Secondary Bus Number Register和Subordinate Bus Number Register的范围内，则该Switch Downstream Port将会产生一笔ACS Violation Error。

在real mode和isolation mode下，除function pooling端口全部使用独立主机域的bdf，无论是global地址访问亦或是local地址访问均可在延用该方案的做法。

### 2、ACS Translation Blocking转换阻塞

该部分在PLDA 提供的第三方IP DN controller中进行处理，当Downstream Port的ACS Extend Capability的ACS Translation Blocking的Enable Bit被使能时，Downstream Port将会检查该Port接收到的Upstream Memory Request的Address Type（AT），如果AT域不是default值(00)该Switch Downstream Port将会报告一个ACS Violation Error。

由于该功能只判断AT字段并不判断地址，所以可延用在controller进行判断处理。

以上两种ACS访问能力在PLDA具体处理如下所示：

**表2 来源验证和转换阻塞处理方式**



### 3、ACS P2P Request Redirect请求事务重定向

访问控制权限是否有效还可能受ACS P2P Egress Control（第6条）和ACS Direct Translated P2P（第7条）访问控制权限影响。当Switch Downstream Port的ACS Extended Capability中的ACS P2P Request Redirect被使能，该Switch Downstream Port接收到P2P Requests必须被向上Redirect到RC。当Root Port的ACS Extended Capability中的ACS P2P Request Redirect被使能，P2P Request必须发送到RC的Redirected Request Validation逻辑单元中，由RC的Redirected Request Validation决定是否将该P2P Request Redirect到原始的Target中，还是被RC block住并产生一笔ACS Violation Error。

根据PCIe协议得如下表1出口控制、请求事务重定向和出口控制向量比特关系映射。本模块根据表3进行映射处理。

**表3 出口控制、 请求事务重定向和出口控制向量比特关系映射**

|  |  |  |  |
| --- | --- | --- | --- |
| E  P2P Egress  Control  出口控制 | R  P2P Request  Redirect  请求事务重定向 | 出口控制  向量比特 acs\_egress  \_ctrl\_vec | 对P2P请求  的处理方式 |
| 0 | 0 | x | 直接路由至P2P目的地 |
| 0 | 1 | x | 上行重定向 |
| 1 | 0 | 1 | 视为ACS违例 |
| 1 | 0 | 0 | 直接路由至P2P目的地 |
| 1 | 1 | 1 | 上行重定向 |
| 1 | 1 | 0 | 直接路由至P2P目的地 |

### 4、ACS P2P Completion Redirect完成事务重定向

当Switch Downstream Port的ACS Extended Capability中的ACS P2P Completion Redirect被使能，P2P的Completions TLP包中的Relaxed Ordering Attribute Bit不为1时，该P2P的Completions TLP必须被向上redirect到RC。否则P2P Completion TLP包将会按照原始目标进行传输。

本模块针对CPLD/CPL类型的报文，如果C标记配置使能且当前报文数据的relax标记为0，则将该CPLD/CPL发送到UP口。

### 5、ACS Upstream Forwarding上行转发

当Switch Downstream Port的ACS Extended Capability中的ACS Upstream Forwarding被使能，Switch Downstream Port的Ingress Port接收到来自下游设备发送的Upstream Request或者Completion TLP（其目标地址是Switch当前的Egress Port，自己发给自己的情况，例如同一个设备的multi\_function互访，或是虚拟地址与物理地址重叠），Switch Downstream Port将接收到的TLP包向上发送到RC。

### 6、ACS P2P Egress Control出口控制

当Switch支持ACS P2P Egress Control时，Switch可以通过选择性配置禁止或者允许某些Switch Downstream Port之间进行P2P传输。系统软件可以通过配置ACS Extended Capability中的Egress Control Vector寄存器来选择让哪些Switch Downstream Ports下的devs之间进行P2P传输。

本模块设计将第3项ACS P2P Request Redirect请求事务重定向和第6项ACS P2P Egress Control出口控制合并处理，处理方式同表2。

### 7、ACS Direct Translated P2P定向转换

当Switch Downstream Port的ACS Direct Translated P2P被使能，AT（Address Type）域的值表示为Translated Address的P2P Memory Requests必须直接被路由到peer Egress Port，无需关注ACS P2P Request Redirect（第3项）和ACS P2P Egress Control（第6项）的使能情况。而对于其他类型的P2P Requests（AT域表示为非Translated Address类型）需要考虑ACS P2P Request Redirect和ACS P2P Egress Control配置。

AT与PCIe总线的地址转换相关，在一些PCIe设备中设置了ATC（Address Translation Cache）部件，该部件的主要功能是进行存储器域和PCI域地址转换，但主要目的是方便多个虚拟主机共享同一个PCIe设备；

1、AT字段为0xb00：TLP的Address字段没有通过ATC进行转换，存放的PCI总线域物理地址；如果PCIe设备进行DMA操作，那么该地址将被RC转化为存储器域物理地址，然后对存储器域进行读写操作；

2、AT字段为0xb01：表示当前TLP报文为“Translation Request”报文，该报文由PCIe设备通过存储器读写请求TLP发出，其目的地为TA；而当TA收到该报文后将根据I/O页表的设置，将合适的地址转换关系，通过存储器读完成TLP，发送给PCIe设备；

3、AT字段为0xb10：表示当前的TLP的Address字段已经通过ATC进行地址转换，当PCIe设备进行DMA操作时，RC不需要进行地址转换，而直接将数据发送给存储器。

### 8、小结

如图2所示，该部分仅处理途中所示左侧分支（即DSP分支）。

首先将支持选项1配置配置到port controller中，进行来源验证，其优先级最高。

将需求支持的6个控制选项按照各个选项的功能和tlp报文type类型进行映射最终确定其路由方向，具体情况如表4所示。

**表4 tlp报文根据type的ACS路由方向**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 分类 | P2P req redirect （控制选项3/6） | | | P2P cpl redirect （控制选项4） | Up forward （控制选项5） | Direct Translated（控制选项7） | 无ACS |
| Type | P2P | 上行  重定向 | 违例 | 上行  重定向 | 上行  重定向 | P2P | P2P |
| MWR | P2P【4】 | UP【5】 | discard【1】 | —— | UP【2】 | P2P【3】 | P2P【6】 |
| MRD/  AOP | P2P【4】 | UP【5】 | cpl\_ca 【1】 | —— | UP【2】 | P2P【3】 | P2P【6】 |
| IO | P2P【3】 | UP【4】 | cpl\_ca 【1】 | —— | UP【2】 | —— | P2P【5】 |
| MES\_ID | P2P【3】 | UP【4】 | discard【1】 | —— | UP【2】 | —— | P2P【5】 |
| CPL | —— | —— | —— | UP【2】 | UP【1】 | —— | P2P【3】 |

**注意：表3中【N】代表该路由在某行报文中的优先级顺序，数值越小优先级越高。其中违例（即Violation）优先级在各种报文类型中均是最高优先级，但是根据第七项的规定，所以当MRD和MWR类型在ACS Direct Translated P2P定向转换使能成立时，需屏蔽违例（Violation）。该部分会影响MRD和MWR类型报文的优先级。**

其中表3中的Type分类的细化详见表4。

表4 ACS涉及Type分类明细表

|  |  |  |  |
| --- | --- | --- | --- |
| Type类 | 具体类别 | Format和Type值 | 报文描述 |
| MWR | MWR64 | 8'h60 | Memory Write为64位地址 |
| MWR32 | 8'h40 | Memory Write为32位地址 |
| MRD | MRD64 | 8'h20 | Memory Read为64位地址 |
| MRD32 | 8'h00 | Memory Read为32位地址 |
| AOP | AOP64\_FA | 8'h6C | AtomicOp FetchAdd为64位地址 |
| AOP32\_FA | 8'h4C | AtomicOp FetchAdd为32位地址 |
| AOP64\_SWAP | 8'h6D | AtomicOp Swap为64位地址 |
| AOP32\_SWAP | 8'h4D | AtomicOp Swap为32位地址 |
| AOP64\_CAS | 8'H6E | AtomicOp CAS为64位地址 |
| AOP32\_CAS | 8'H4E | AtomicOp CAS为32位地址 |
| IO | IORD | 8'h02 | IO read |
| IOWR | 8'h42 | IO write |
| MES\_ID | MES\_ID | 8'h32 | Message无data |
| MESD\_ID | 8'h72 | Message有data |
| CPL | CPL | 8'h0A | Cpl无data |
| CPLD | 8'h4A | Cpl有data |

## function pooling端口方案

### 1、拓扑结构映射



图4 function pooling拓扑结构第一次映射

如图4所示，在function pooling端口下，由于多个主机域共享同个PCIe树结构（当前称之为global域），无法确定各独立主机域（local域）的ACS配置不同配置，所以要求所有报文全部上行重定向到function pooling端口后，根据各个域的每一级ACS配置进行解析。

考虑不限制用户在实际接PCIe树的行为模式，采用将多层级映射为固定虚拟两级的拓扑结构方式。按照用户对每个实际EP的function切分，实际将切割后的function独立为主机域的EP。

当前方案的做法先进行第一次映射，按照深度优先的规则将所有切割后的EP，每32个EP按照顺序下挂在第二级SW DP下，系统将限制总EP的个数MAX\_EP（以下用该值为512举例）。

基于以上的拓扑结构映射，所有的EP将根据function进行拆分，其ID和ADDR均全部保持global域的值进行划分，且保持深度优先的规则，那么每个EP都可通过bus number和function number查找到唯一的EP，可以推出使用将虚拟SW中的ID路由比较由原始的bus num是否在secondary bus num和subordinate bus num范围内，转换为{bus num，function num}是否在{secondary bus num，function num}和{subordinate bus num，function}范围内。此时的function number取决于用户配置和32个function拆分的EP重组的情况。

同理，ADDR拆分和重组的方式与ID的方式完全相同，该部分目前定义有芯片内部的MCPU进行处理。

由以上可以看出function pooling端口拓扑结构第一次映射，ADDR域和ID域全部不变保持global域，仅重新根据function划分EP将拓扑结构重组，MCPU需要将中间虚拟层级SW的ID范围和ADDR范围进行重新计算，以方便后续其他处理需要。



图5 function pooling拓扑结构第二次映射

将虚拟两层结构映射完成后，按照第一次映射拆分后的EP属于主机域的归属，各自呈现各主机域独立的拓扑结构，如图5所示。

定义主要规则如下：

1. 上电后实际接的多层拓扑结构按照用户对function的切分，直接进行两层拓扑结构的映射。
2. 正常运行后的原结构EP不动，仅进行function的动态切换，那么第一次映射的拓扑结构不变，仅第二次拓扑结构给不同的主机域呈现热插拔操作，即原有主机域拔出新主机域插入操作。需要注意的为插入的位置，由于硬件处理ACS并不需要主机域地址进行映射，该部分插入操作由系统和软件定义，建议其在主机域未使用的bus number最后预留位置上线。
3. 当前暂不支持实际接的EP因为热插拔出现新增加function映射功能。

### 2、ACS配置

该部分为MCPU为function pooling端口下挂的真实端口的基于ACS功能配置空间的配置定义，要求必须支持且打开ACS能力，具体分类如下：

1、ACS Source Validation来源验证

开启来源验证功能，该部分要求系统针对不同主机域具有相同的处理方式，则虚拟两级可不再重复校验该部分。

2、ACS Translation Blocking转换阻塞

关闭转换阻塞功能，该部分由虚拟两级逐层判断。

3、ACS P2P Request Redirect请求事务重定向

开启请求事务重定向，关闭出口控制，如表3所示，要求所有请求事务全部上行重定向。

4、ACS P2P Completion Redirect完成事务重定向

开启完成事务重定向，并通过业务标记Relaxed Ordering Attribute Bit为0（需商榷）。

5、ACS Upstream Forwarding上行转发

开启上行转发。

6、ACS P2P Egress Control出口控制

关闭出口控制。

7、ACS Dirext Translated P2P定向转换

关闭定向转换功能。

