基于细粒度访问控制的大数据安全防护方法

王继业,范 永,余文豪,韩丽芳
(中国电力科学研究院有限公司,北京 100192)

摘 要: 访问控制是保护信息系统数据安全的重要手段。但是大数据服务环境下,数据呈现分布式的特点。如何有效解决复杂用户多数据资源域的访问,是大数据安全的重要研究方向。针对这一问题,在深入分析大数据环境下访问控制技术的基础上,提出了一种基于细粒度访问控制的大数据安全防护方法。该方法采用基于属性的访问控制模型,解决了用户认证、域定位、访问决策以及模块关联的问题,实现了细粒度数据及服务的访问。在提出基本模型之上,结合实际的应用场景需求,给出了单域和跨域两个场景中的访问决策模型。详细描述了模型及决策算法,并给出了多域属性表同步方法。实验结果表明,该模型实现了细粒度访问,能够有效保护大数据环境下的数据安全,并且能够实现快速决策,高效访问。

关键词: 大数据; 信息安全; 访问控制; 属性; 细粒度
中图分类号: TP309 **文献标识码**: A **文章编号**: 1673-629X(2019)10-0134-07
doi: 10.3969/j.issn.1673-629X.2019.10.027

Big Data Security Protection Based on Fine-grained Access Control

WANG Ji-ye,FAN Yong,YU Wen-hao,HAN Li-fang
(China Electric Power Research Institute,Beijing 100192,China)

Abstract: Access control is an important way to protect data security of information system. However,in the context of big data services, data is distributed. How to effectively solve the access of multiple data resource domains of complex users is an important research direction of big data security. Aiming at this problem,we put forward a big data security protection based on fine-grained access control by the in-depth analysis of access control technology in big data environment. The method adopts an attribute-based access control model to solve the problems of user authentication,domain location,access decision and module association,and achieves fine-grained data and service access. On the basis of the proposed basic model,combined with the actual application scenario requirements,the access decision model in the single domain and cross-domain scenarios is given. We describe the model and decision algorithm in detail,and give a multi-domain attribute table synchronization method. The experiment shows that the model can achieve fine-grained access,effectively protect data security in big data environment,and can achieve fast decision making and efficient access.

Key words: big data; information security; access control; attribute; fine-grained

0 引 言

随着信息技术和网络技术的快速变革,以及物联网、云计算、人工智能等新兴技术的发展与应用,大体量、丰富结构、复杂类型、智能分析的大数据服务时代已经来临[1-4]。

但是大数据服务时代在数据安全上也存在很大的隐患和挑战[5]。大数据安全问题主要体现在 3 个方面: 第一是信息泄密[6]。开放的互联网使海量的大数据处于裸奔状态,国内网络设备主要依赖国外技术,很容易因为“漏洞”、“后门”造成信息泄露。第二是数据非授权访问[7]。主要是复杂关系的多用户存储在云盘、大数据平台等共享平台的数据所有权归属及访问问题。第三是网络攻击[8]。大数据服务时代的大规模数据存储,只提供单一的访问服务接口,易遭受黑客的大规模网络攻击,进而影响大规模用户的正常数据使用。

大数据服务采用分布式方法存储数据,并基于大数据平台将海量的数据资源链接起来,构建大规模的数据开放共享平台。所有位于大数据环境中的数据所有者,将数据存放于大数据平台进行统一管理,进而构

收稿日期: 2018-11-05 **修回日期**: 2019-03-06 **网络出版时间**: 2019-04-24
基金项目: 国家电网公司总部科技项目(JS71-16-005)
作者简介: 王继业(1964-),男,博士,高级工程师(教授级),研究方向为电力系统信息安全、大数据等。
网络出版地址: http://kns.cnki.net/kcms/detail/61.1450.TP.20190424.1051.058.html

成一个巨大的数据服务中心。大数据中心不仅能够为数据所有者提供存储和访问服务,又能够支持各种复杂的数据运算。在大数据服务环境下,传统架构模式下的单一数据中心消失,而是以分布式数据存储的形式存在。数据资源的存储分配从单一的主机存储分配变为多主机、分布式的存储分配。数据资源的访问也从单一用户单一主机访问变为多用户多数据域的数据访问,从而引出了大数据服务环境下的复杂用户多数据资源域的访问问题。访问控制模型的构建是解决数据访问的重要方法。因此,在深入研究传统访问控制模型的基础上,文中提出了面向大数据服务环境下的基于细粒度访问控制模型的安全防护方法。

1　相关研究

传统访问控制模型分为三个主要类型: 自主访问控制模型、强制访问控制模型和基于角色的访问控制模型。传统的访问控制模型通常是基于表示或者特定的身份,对数据资源进行分组的访问管理。但是这种方式存在局限性,只能用于结构简单的封闭式网络环境。由于大数据环境面临用户数量多且用户关系复杂的问题,数据资源需求存在很多不确定性,权限的授权与取消必须是动态变化的。更复杂的是,用户的访问请求可能涉及不同逻辑区域的分布式数据存储资源。由此可见,传统的访问控制模型不适用于分布式下的大数据环境。

当前,国内研究者针对大数据环境的访问控制需求的特点,提出了很多访问控制模型,其本质上是对传统访问控制模型进行的扩展和改造。

崔新会等[9] 对大数据环境给访问控制领域带来的问题进行了系统分析,提出了 5 个迫切问题: 授权管理、细粒度访问控制、访问控制策略描述、个人隐私等,并给出了若干建议,但没有进行实质性的研究。惠榛等[10] 以大数据医疗为背景,提供了一种基于风险的访问控制模型,能够适应性地调整医生的访问能力,保护患者隐私。但是此方法需要分析医生的访问历史来量化风险,推断访问权限,对数据质量及环境要求较高,不能适用于统一化、标准化的大数据处理平台。王小明等[11] 针对大数据的强动态性及强隐私性,笼统地给出了基于属性访问控制模型各个阶段的研究点及解决方法,没有提出具体的模型。路艳军等[12] 提供了基于代理技术的大数据平台的访问控制方法,实现了代理式的数据资源的统一访问接口,但是没有给出细粒度的访问策略。胡坤等[13] 提出了大数据安全和隐私保护风险,给出了大数据安全防护策略,但是没有提出具体的防护方法。闫玺玺等[14] 将属性加密技术和控制技术相结合,提出了针对敏感数据的融合访问控制机制,通过加密技术提高了策略分发的安全性,但是控制访问只是粗粒度的访问策略。

基于国内大数据安全防护方法的调研与研究,文中提出了基于细粒度访问控制的大数据安全防护方法,构建了基于属性的细粒度访问控制模型。在该模型中,访问请求被表述为不同的属性信息,如访问者、访问资源、访问动作、环境等。这些属性同时作为访问依据,能够保证实现动态、细粒度的授权机制,保证访问控制的灵活与可扩展。

2　BD-ABAC 模型

大数据环境下的细粒度属性访问控制模型(big data-attribute based access control,BD-ABAC) 是通过基于属性的细粒度访问控制方法,利用多域属性表同步技术,实现多域细粒度访问控制的模型。

传统的访问控制方法是直接在主体和客体之间定义授权,基于属性的细粒度访问控制方法是利用主体、资源、环境和动作之间的属性关系定义授权。BD-ABAC 模型和基于角色的访问控制模型存在本质上的不同。它将操作与资源之间的二元关系扩展到主体、资源、环境和动作之间的多元约束关系,形成了基于多元属性的细粒度访问控制。

属性是人类对事物的性质与关系的定义,即对现实世界实体的抽象描述。两个实体是否相同,其本质就是实体的属性是否相同。通常把访问控制的需求归结为四种属性: 主体属性、资源属性、环境属性和动作属性。

BD-ABAC 模型能够进行基于属性的细粒度授权决策。为了实现这一目标,BD-ABAC 定义了一种细粒度的灵活授权方法,能够适应大数据环境下的资源跨域访问、实体属性动态变化等特点。下面给出 BD-ABAC 模型相关的数据定义:

(1) 属性定义。

attr 表示属性(attribute) ,attr_ n 表示第 n 个属性。Sattr,Rattr,Eattr 和 Aattr 表示主体属性(subject attribute) 、资源属性(resource attribute) 、环境属性(environment attribute) 和动作属性(action attribute) 。

$\mathrm{Sattrs}_a = \{\mathrm{Sattr\_1}, \mathrm{Sattr\_2}, \cdots, \mathrm{Sattr\_}a\}$,为主体属性集合, a 表示属性个数;

$\mathrm{Rattrs}_b = \{\mathrm{Rattr\_1}, \mathrm{Rattr\_2}, \cdots, \mathrm{Rattr\_}b\}$,为资源属性集合, b 表示属性个数;

$\mathrm{Eattrs}_c = \{\mathrm{Eattr\_1}, \mathrm{Eattr\_2}, \cdots, \mathrm{Eattr\_}c\}$,为环境属性集合, c 表示属性个数;

$\mathrm{Aattrs}_d = \{\mathrm{Aattr\_1}, \mathrm{Aattr\_2}, \cdots, \mathrm{Aattr\_}d\}$,为动作属性集合, d 表示属性个数。

(2) 属性赋值定义。

AV 表示属性赋值(attribute value) 。AV_n ←(attr_n = value) ,表示第n个属性的属性赋值。SAV_n, RAV_n ,EAV_n 和 AAV_n 表示第 n 个主体、资源、环境和动作属性赋值。

SAV_t = { SAV_1,SAV_2,…,SAV_t } ,为主体属性赋值集合, t 表示属性赋值个数;

RAV_m = { RAV_1,RAV_2,…,RAV_m } ,为资源属性赋值集合, m 表示属性赋值个数;

EAV_n = { EAV_1,EAV_2,…,EAV_n } ,为环境属性赋值集合, n 表示属性赋值个数;

AAV_k = { AAV_1,AAV_2,…,AAV_k } ,为动作属性赋值集合, k 表示属性赋值个数。

(3) 属性阈值定义。

ATV 表示属性阈值(attribute threshold value) 。ATV_n ←(attr_n $\propto$ value) ,表示第 n 个属性的属性阈值,其中 $\propto \in \{>,<,=,\geqslant,\leqslant,\neq$,in,not in,between} 为关系表达式运算符,用来限定属性的取值范围,即属性关系。SATV_n ,RATV_n ,EATV_n 和 AATV_n 表示第 n 个主体、资源、环境和动作的属性阈值。

$SATV_i$ = { SATV_1,SATV_2,…,SATV_i } ,为主体属性阈值集合,其中 i 表示属性阈值个数;

$RATV_j$ = { RATV_1,RATV_2,…,RATV_j } ,为资源属性阈值集合,其中 j 表示属性阈值个数;

$EATV_s$ = { EATV_1,EATV_2,…,EATV_s } ,为环境属性阈值集合,其中 s 表示属性阈值个数;

$AATV_p$ = { AATV_1,AATV_2,…,AATV_p } ,为动作属性阈值集合,其中 p 表示属性阈值个数。

(4) 访问请求定义。

Req(SAV_t,RAV_m,EAV_n,AAV_k) = { SAV_1,SAV_2,…,SAV_t } ∩ { RAV_1,RAV_2,…,RAV_m } ∩ { EAV_1,EAV_2,…,EAV_n } ∩ { AAV_1,AAV_2,…,AAV_k } ,表示一个完整的由主体、资源、环境和动作属性赋值构成的访问请求。

(5) 访问控制策略定义。

Policy ←(Target, CombiningAlgorithm, Rule) ,表示访问控制策略,由 Target (目标)、CombiningAlgorithem(合并算法) ,Rule(规则) 三者组成。

Target←($Sattrs_a$, $Rattrs_b$, $Eattrs_c$, $Aattrs_d$) ,表示目标,由主体属性集合(Sattrs) ,资源属性集合(Rattrs) ,环境属性集合(Eattrs) 和动作属性集合(Aattrs) 组成,用来说明所属的访问控制策略(Policy) 是否适合访问请求。

Rule = { $rule_1$, $rule_2$,…, $rule_t$ } ,表示规则集合,其中 rulet 表示第 t 条规则。rule = Result←($SATV_i$, $RATV_j$, $EATV_s$, $AATV_p$) ,表示一条完整的规则组成。其中 Result 为规则的判定结果,Result ∈ (Yes,No) 。

CombiningAlgorithem(合并算法) ,表示生成决策结果对规则(Rule) 判定结果(Result) 的解决策略冲突的合并算法。

2.1 模型构建

BD-ABAC 模型(见图 1) 由三个核心模块组成:用户认证模块、域定位模块和访问决策模块。

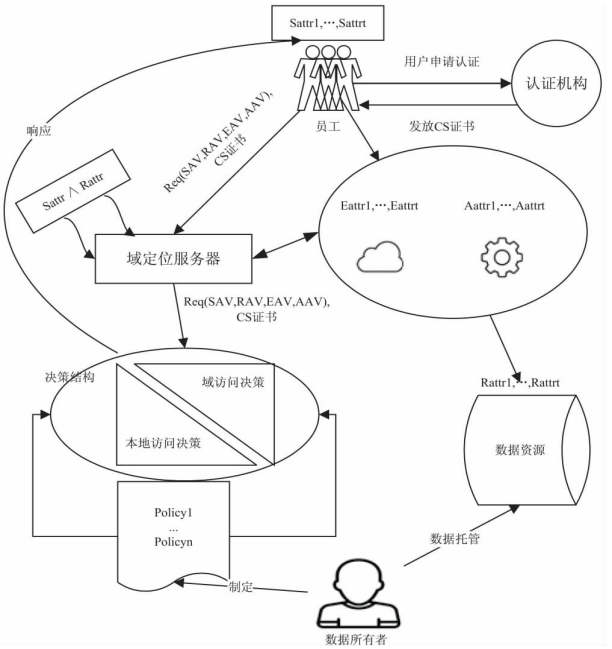

图 1 BD-ABAC 模型

2.1.1 用户认证模块

用户在第一次访问数据资源的时候,首先要向安全认证机构申请安全证书(security certificate, SC) 。安全证书的作用在于,用户可以使用安全证书在访问数据资源时证明自己的身份。资源服务器的作用是对抗恶意的数据访问,验证的凭据也是身份证书。这样设计的好处在于,安全证书被全域信任,能够有效减少重复的验证操作。

2.1.2 域定位模块

通过身份认证的终端用户在访问数据资源时,访问请求 Req(SAV_t,RAV_m,EAV_n,AAV_k) 会被转发至域定位服务器,域定位服务器对主体属性(Sattr) 和资源属性(Rattr) 进行分析,判断该资源的访问请求是本地域访问还是跨域访问。如果是跨域访问请求,域定位模块则根据资源属性(Rattr) 快速查找相应的资源域,并转发访问请求。

2.1.3 访问决策模块

访问决策模块由两部分组成,分别是本地域访问决策部分和跨域访问决策部分。采用基于属性的访问控制方法对终端用户的访问请求 Req(SAV_t,RAV_m,EAV_n,AAV_k) 进行策略判定,并将判定结果发送到策略执行点。各数据资源域根据属性策略集合相应的合

并算法,为策略判断提供可靠依据。

2.1.4　模块关联

在 BD-ABAC 模型中,当终端用户申请访问数据资源时,用户认证模块、域定位模块和访问决策模块之间的交互协作关系如图 2 所示。终端用户向数据资源发起访问请求,通过使用安全认证机构颁发的安全证书进行身份认证,通过身份认证的终端用户的访问请求被转发至域定位服务器,域定位服务器将访问请求转发至相应的资源域,该资源域的决策机构对该访问请求进行判定并将结果返回至终端用户。

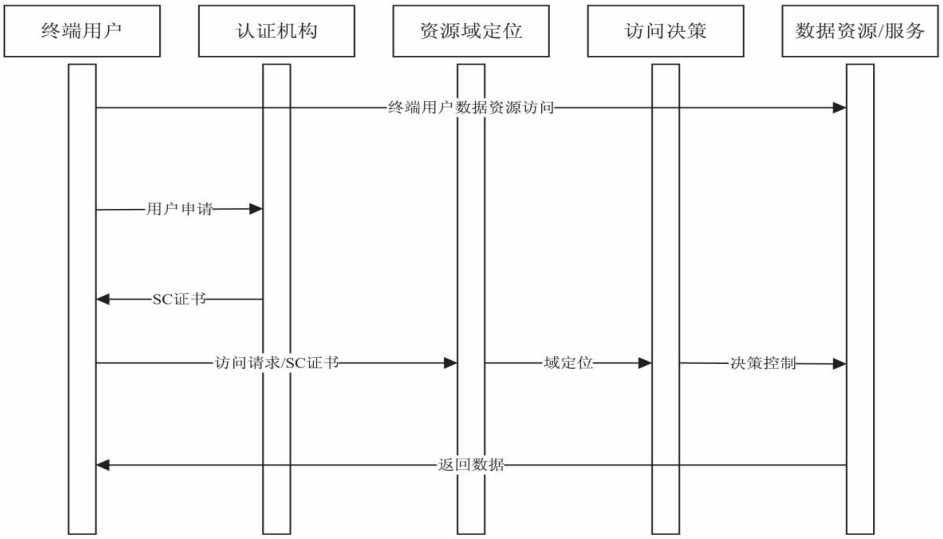

图 2　模块协作

2.2　单域访问决策模型

文中采用可扩展访问控制标记语言(XACML) [15] 框架来制定单域访问控制的决策模型。XACML 是一种标准化的框架,提供了统一的访问控制策略编写规范,能够支持大数据计算环境下的多数据域之间的访问控制。

在单域访问的环境中,访问控制决策模型由四个主要部分组成: 策略执行点(PEP) 、策略决策点(PDP) 、策略信息点(PIP) 、策略管理点(PAP) 。其中,PEP 负责将用户的访问请求传递到 PDP,并得到决策结果。PDP 负责对访问请求进行判断,并把判定结果返回给 PEP。PIP 的作用是为 PDP 决策提供属性信息。PAP 为 PDP 进行决策提供所必需的策略集。

文中提出的单域访问决策模型如图 3 所示。单域访问策略判定过程如下:

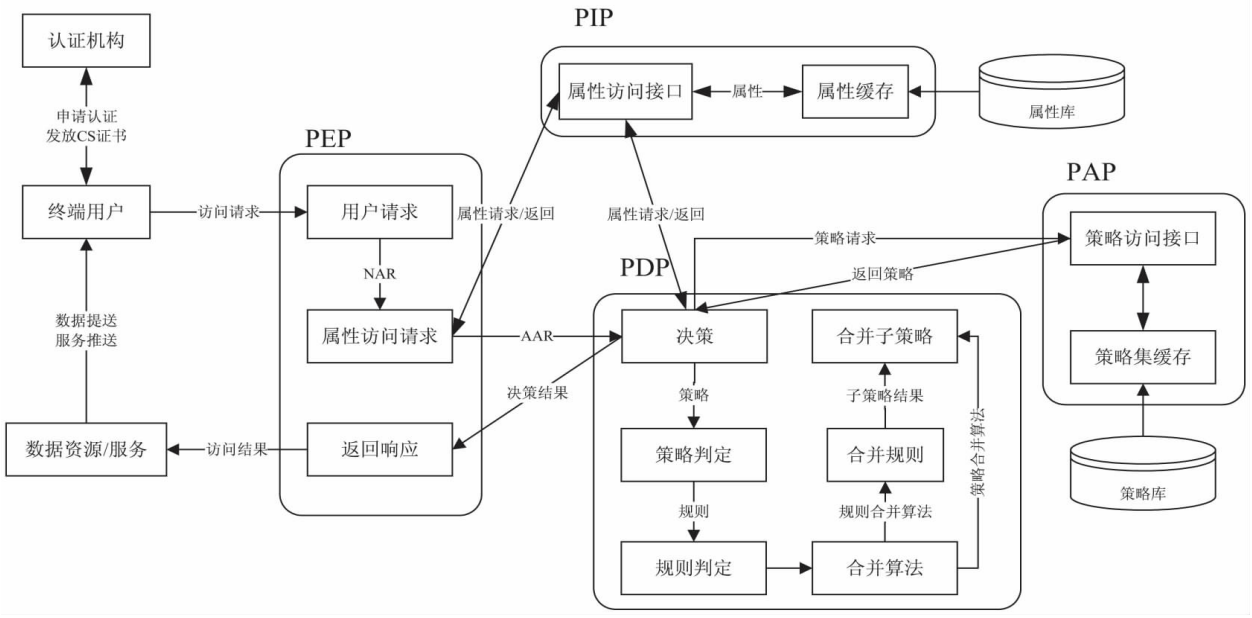

图 3　单域访问决策模型

(1) 用户第一次访问数据资源,进行身份认证,并申请用户安全证书 SC。

(2) 用户发送访问请求到域定位服务器,域定位服务器通过对主体和资源的属性分析,判断为单域访

问服务。

(3) PEP 接收到访问请求 NAR 后,根据访问请求的内容,结合 PIP 和属性库,构建基于属性的访问请求 AAR。AAR 包含了本次访问所需的所有相关属性。然后,PEP 将 AAR 转发给 PDP。

(4) PDP 负责对访问请求进行决策。为了查找适合该 AAR 的策略集,PDP 根据 AAR 和 PIP 中的属性,向 PAP 发送策略匹配请求,进而得到结果。

(5) PAP 根据 AAR 中的属性在策略库中查找适合的策略集,并返回给 PDP 进行策略判定。

(6) 策略集通常包含了多个策略,而每条策略又由多种规则组成,而这些规则元素就是判断依据。

(7) 策略合并算法为多条策略决策结果进行合并的方法。

(8) PDP 将判定结果返回给 PEP。此时,PEP 能够根据判定结果执行用户访问数据资源的决策。

单域访问决策中最重要的模块是 PDP 模块,PDP 模块的功能是: 接收到由 PEP 转发来的基于属性的访问请求后,通过 PIP 和 PAP 获取属性信息和策略集,对请求进行策略判定,并返回判定结果。

2.3 跨域访问决策模型

在当今的复杂应用背景下,大数据环境通常存在多个不同领域,例如金融、安防、能源、业务、医疗以及电力行业等。各领域中通常存在有自己的安全认证和访问控制体系,终端用户在访问不同领域的大数据资源时,如何保证用户安全高效的在多域之间进行互操作是访问控制模型必须考虑的。文中在提出基于属性的访问控制模型的基础上,对大数据环境下多域间安全访问以及资源的互操作提出解决方案。访问控制模型如图 4 所示。

(1) 位于主体域的用户,在经过安全认证的前提下,持有安全证书进行跨域访问。

(2) 在访问请求抵达资源域时,首先要通过策略决策,主体域判断用户是否有权限去访问资源域,具体由主体域的策略判定点进行决策。

(3) 主体域的策略判定点将决策结果发送到主体域的策略执行点和主体域的策略执行点执行主体域的策略判定点决策结果中的权限。

(4) 主体域的策略执行点发送多域间访问请求到资源域的策略执行点,资源域的策略执行点查看其安全证书。

(5) 资源域的策略执行点首先检查主体属性是否发生了变化。若没有变化,就发起属性库更新请求,更新资源域属性库中主体的属性表信息。若没有变化,则将请求提交到下一步。

(6) 资源域的策略执行点发送访问请求到资源域的策略判定点和资源域的策略判定点匹配适用于此访问请求的策略集进行策略决策。

(7) 资源域的策略判定点将最终的决策结果发送到资源域的策略执行点。

(8) 资源域的策略执行点执行策略判定点决策结果中的权限,对用户的访问请求进行判决,允许或拒绝资源的访问请求。

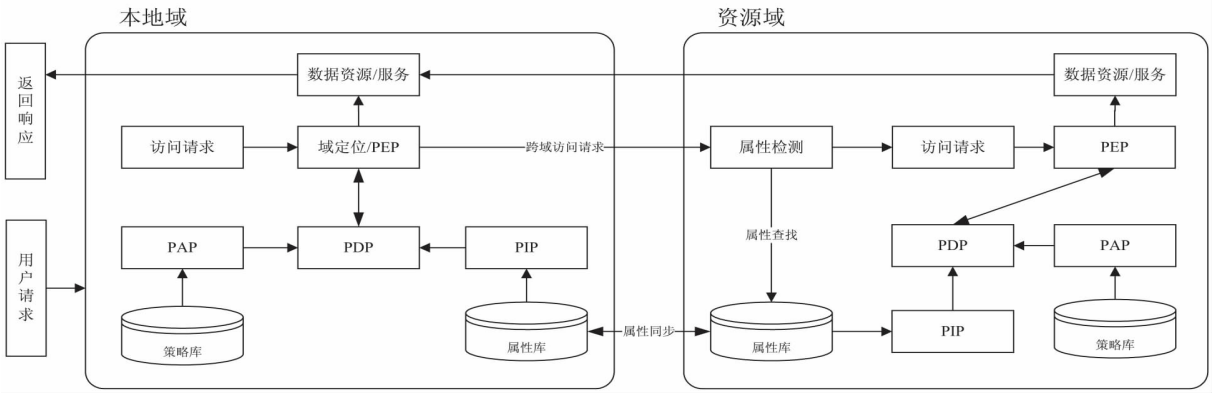

图 4 跨域访问决策结构模型

2.3.1 跨域属性表同步

尤其是在大数据环境下,用户权限是动态变化的,主体信息经常发生变化会影响策略的评估和判定,各操作域中的访问控制模型需要一种可靠机制能够及时更新属性库中的信息。属性表的更新需要重点解决以下问题:

(1) 在多域间互操作的过程中,当本地域属性表发生变化时如何保持其在本地域和资源域的一致。

(2) 当对资源域的属性表进行更新时,如果有其他操作调用此属性表中的信息,则有可能产生误操作。

针对上述问题,为了解决多域属性表更新同步互斥的问题,文中引入信号量机制和 P/V 操作。属性库中的每一个主体都赋予一个信号量,并设置初始值为 1: mutexj = 1。当策略执行点 PEP 发现属性变化时,发送更新请求,执行 P 操作。更新完毕后,执行 V 操作。执行完毕后,将属性更新结果返回主体域。P/V 操作

过程如图 5 所示。

```
Produce P(Var mutex:Semaphore):
Begin
 mutex:=mutex-1;
 if mutex<0
  then
   //主体J的属性表被其他操作调用无法进行更新操作
   W(mutex);
  else
   Update(J.attribute list);
End;
Produce V(Var mutex:Semaphore):
Begin
 mutex:=mutex+1;
 if mutex<=0
  then
   //从等待队列中唤醒正在等待的调用主体J属性的操作
   R(mutex);
End;
```

图 5　P/V 操作过程

资源域的策略执行点的主要执行流程为: 检测访问请求中属性表是否发生变化,若发生变化则执行更新同步属性表的操作,若未发生变化则将访问请求传递给策略判定点。

2.3.2　访问决策核心算法

假设访问请求为 Req(SAV_t, RAV_m, EAV_n, AAV_k) ,策略为 Policy←(Target,CombiningAlgorithm, Rule) ,首先进行策略评估 PolicyEvalucation(Req,Policy) 。含义是根据访问请求和策略,决定访问请求是否能够被批准。若此策略适合判定,则返回策略的 Result 集合,否则不返回。策略判定算法如图 6 所示。

```
//评估策略是否适合判定访问请求
PolicyEvalucation(Req(SAV，RAV，EAV，AAV), Policy)
{
  //判断策略属性是否包含请求属性
  if(SAV.Sattr∈Policy.Target.Sattrs &&
     RAV.Rattr∈Policy.Target.Rattrs &&
     EAV.Eattr∈Policy.Target.Eattrs &&
     AAV.Aattr∈Policy.Target.Aattrs)
    //策略判定
    PolicyDecition(Req(SAV, RAV, EAV, AAV), Policy.Rule);
  else
    return null;
}

//判定请求是否满足策略规则
PolicyDecition(SAV, RAV, EAV, AAV), Rule[])
{
  //判定规则结果
  for(i=0;i<Rule.length;i++)
  {
    if(Req.AV.attr==Rule[i].ATV.attr &&
       Req.AV.value∝Rule[i].ATV.value)
    {
      results[i]=Rule[i].Result;
    }
  }
  //合并规则集的判定结果
  result= Combining(results[],CombiningAlgorithem);
  //返回策略判定结果
  return result;
}
```

图 6　策略判定算法

在多域访问的场景中,保持主体域和资源域中的属性一致性是至关重要的。资源域中的 PEP 模块的主要功能是: 检查访问中的主体属性和先前属性是否一致,如果没有变化,就交送 PDP 进行进一步的访问决策。如果发生了改变,就要更新属性库中的属性表。PEP 模块核心算法如图 7 所示。

```
PEP.CoreAlgorithm(Req(s,r,e,a))
Begin
 //定义默认标记值
 flag=1;
 //对访问请求中主体进行属性检测
 flag=Detection(Req.s.attr,s.AttrTable)
 //检测到属性改变
 if(flag==0)do
 Begin
  P(s.mutex)
  //更新主体属性表
  Update(s.AttrTable)
  V(s.mutex)
 End
 PDP.coreAlgorithm(Req(s,r,e,a))
End
```

图 7　PEP 模块核心算法

3　仿真实验

实验基于 Hadoop 大数据处理平台[16-17],在该平台上模拟 3 个数据资源存储中心,并划分为 3 个不同的逻辑域,以此命名为 M1,M2 和 M3,如图 8 所示。

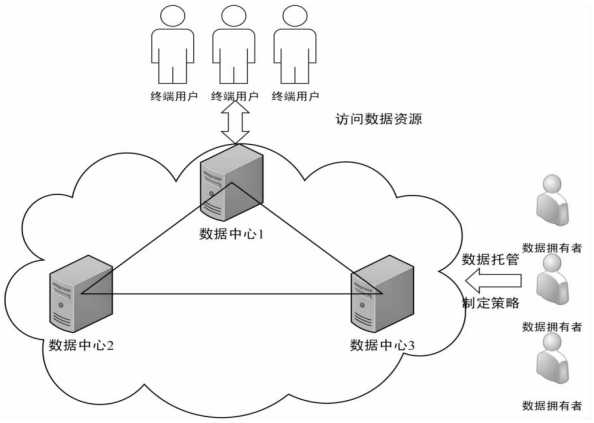

图 8　大数据分布式处理平台

在三个逻辑域中建立多个仿真用户,用户角色包括数据拥有者和数据用户。数据拥有者在不同的逻辑域中掌控不同的数据资源。并且进一步根据属性对权限进行细粒度的定义,制定完备的访问侧逻辑和,并对满足属性约束的用户授予访问控制权限。

实验设置如下: 针对 3 个逻辑域,模拟编写 1 000 条策略,其中每条策略包含 1 条至上千条规则不等,以验证 BD-ABAC 模型的正确性、细粒度访问决策、可扩展性和效率等。

3.1　细粒度访问控制

访问控制策略复杂度和访问控制粒度相关,越复杂代表控制粒度越细。策略中包含的属性个数越多,策略的复杂度就越高。实验测试中选择了 50 条访问控制策略,这些策略的复杂度不同。并且每条策略含有多个属性阈值对,从几个到几千不等。实验测试结

果如图 9 所示,记录了策略复杂度与策略决策时间的关系。

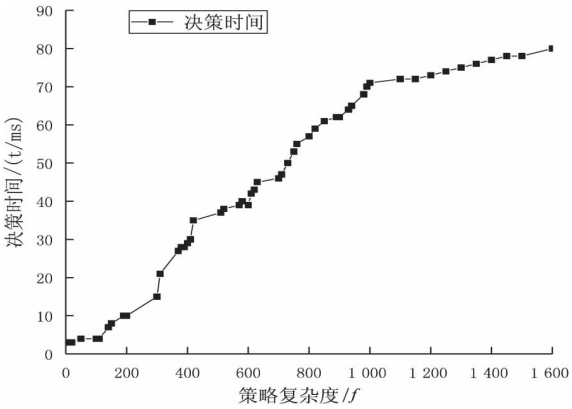

图 9 策略复杂度与策略响应时间的关系曲线

3.2 单域及多域间访问决策

在单域访问控制决策和多域间访问控制决策的实验测试中,随机选择了二十条请求,并记录策略判断的决策时间,重点分析请求与策略规则数目的关系。重复五十次实验取平均值。图 10 为不同规则数与策略决策时间的关系曲线。

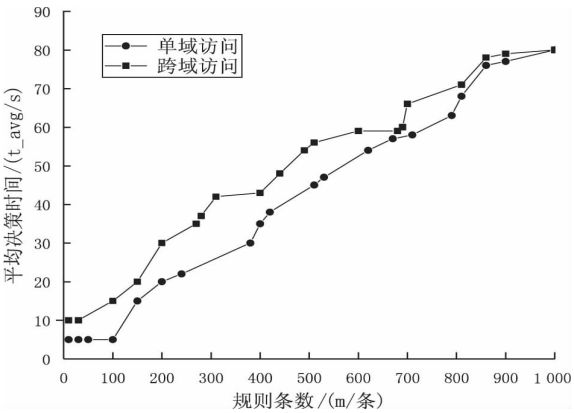

图 10 策略条数与平均决策时间的关系曲线

由图 10 可知,无论是单域还是多域间访问请求,其策略判定的平均决策时间都随访问控制策略规则数的增加而上升,并趋近于平缓。而多域间访问平均决策时间相对更长,但时间开销在可接受范围内。在实际情况下,策略条数并没有实验中这么多,策略也不会这么复杂,所以,判断时间不会影响系统的性能。综上所述,BD-ABAC 具有较好的可扩展性及较高的决策效率。

4 结束语

BD-ABAC 结合安全认证和访问判决两个维度,使授权验证的安全性和可信度更高,这有助于提高大数据资源访问的安全性。在访问控制粒度方面,BD-ABAC 基于多种角色的属性进行判决,采用了动态灵活细粒度的访问授权机制,并引入了信号量和 P/V 操作有效解决了跨域更新互斥问题。实验结果表明,BD-ABAC 模型正确,可扩展,能够满足大数据资源访问高效和细粒度的要求。

参考文献:

[1] 李国杰,程学旗.大数据研究: 未来科技及经济社会发展的重大战略领域—大数据的研究现状与科学思考[J].中国科学院院刊,2012,27(6):647-657.

[2] ZIKOPOULOS P,EATON C.Understanding big data: analytics for enterprise class Hadoop and streaming data[M].[s.l.]: McGraw-Hill Osborne Media,2011.

[3] BOYD D,CRAWFORD K.Critical questions for big data: provocations for a cultural,technological,and scholarly phenomenon[J].Information,Communication & Society,2012,15(5):662-679.

[4] LABRINIDIS A,JAGADISH H V.Challenges and opportunities with big data[J].Proceedings of the VLDB Endowment,2012,5(12):2032-2033.

[5] 冯登国,张 敏,李 昊.大数据安全与隐私保护[J].计算机学报,2014,37(1):246-258.

[6] 程学旗,靳小龙,王元卓,等.大数据系统和分析技术综述[J].软件学报,2014,25(9):1889-1908.

[7] 孟小峰,慈 祥.大数据管理: 概念、技术与挑战[J].计算机研究与发展,2013,50(1):146-169.

[8] 方滨兴,贾 焰,李爱平,等.大数据隐私保护技术综述[J].大数据,2016,2(1):1-18.

[9] 崔新会,陈 刚,何志强.大数据环境下云数据的访问控制技术研究[J].现代电子技术,2016,39(15):67-69.

[10] 惠 榛,李 昊,张 敏,等.面向医疗大数据的风险自适应的访问控制模型[J].通信学报,2015,36(12):190-199.

[11] 王小明,付 红,张立臣.基于属性的访问控制研究进展[J].电子学报,2010,38(7):1660-1667.

[12] 陆艳军,李月航,李忠强.大数据平台访问控制方法的设计与实现[J].信息安全研究,2016,2(10):926-930.

[13] 胡 坤,刘 镝,刘明辉.大数据的安全理解及应对策略研究[J].电信科学,2014,30(2):112-117.

[14] 闫玺玺,耿 涛.面向敏感数据共享环境下的融合访问控制机制[J].通信学报,2014,35(8):71-77.

[15] LORCH M,PROCTOR S,LEPRO R,et al.First experiences using XACML for access control in distributed systems[C]//Proceedings of the 2003 ACM workshop on XML security.Fairfax,VA,USA: ACM,2003: 25-37.

[16] SHVACHKO K,KUANG H,RADIA S,et al.The Hadoop distributed file system[C]//IEEE symposium on MASS storage systems and technologies.Incline Village,NV,USA: IEEE,2010: 1-10.

[17] BORTHAKUR D,GRAY J,SARMA J S,et al.Apache Hadoop goes realtime at Facebook[C]//ACM SIGMOD international conference on management of data.Athens,Greece: ACM,2011: 1071-1080.