[智能配电]

面向云存储的智能配电网监测数据完整性验证方法研究

文/博宇融通（北京）电气设备有限公司　姜维权

针对智能配电网云存储数据完整性验证问题中计算开销较大，无法支持海量监测数据存储，且无法支持动态更新的问题，提出了一种面向云存储支持动态更新的数据完整性验证方法。该方法是基于同态标签和Hadoop平台相结合，支持第三方审计、动态更新、海量数据存储，并且计算开销较小。并通过安全性和正确性证明了此方案的有效性。

智能配电网自动化、信息化和互动性的提高，及物联网的相互融合与渗透，造成电力企业的量测体系结构中积累了大量数据，如用户的用电数据、调度运行数据和设备检测数据等。除了内部数据，电力系统还积累了大量运营数据，如客户服务数据和电力市场数据等。这些海量数据在系统决策中发挥着重大的作用，但是如何高效地存储和保护海量智能配电网数据的完整性存储是制约智能配电网发展的关键。

有的文献提出了可证明的数据持有方法，主要是使用同态标签和基于RSA的模指运算进行验证，随后在文献中将PDP方案扩展到支持可公开校验方案，把原来的同态标签改为支持公开验证标签，只要拥有公钥的第三方便可进行数据完整性验证。文献提出了一种基于“哨兵”的云存储数据完整性验证协议，该协议支持无限次检测，通讯开销、本地存储开销和计算开销少，但无可信第三方进行验证；文献提出一种基于Hash算法的云数据完整性验证机制，通过随机抽样数据块的方法计算Hash值，从而有效降低了服务器的开销，且该机制系统开销小，置信度高且支持可信第三方。文献提出了一种云存储数据完整性检测架构，实现了数据动态性和公开可验证性，且支持第三方检测，但计算开销和通信开销较大。

根据上述分析可知，现有的文献对配电网数据完整性已提出了一些检测方法，但是存在一些不足，受到某些文献的启发，本文基于同态标签提出一种智能配电网云存储数据完整性验证方法。

由于云存储开销低、高可用性和高可靠性的优点，本文利用云存储进行海量数据的存储工作。该方案能实现以后目标：

（1）支持动态更新，用户对存储在云端的数据可以进行动态的插入、删除和修改，操作完成后，可以进行数据完整性验证。

（2）公开可验证性，本文验证方法支持第三方进行数据完整性验证，并实现密钥生成服务、文件上传操作、文件预处理操作和云存储数据完整性验证。

（3）计算开销较小，并且所有的服务都在云存储上。

相关技术

1．同态标签

同态是指从一个代数结构映射到另一个代数结构，并保持所有相关的结构不变，即，ϕ：x→

y，满足$\phi(xy)=\phi(x)^{\circ}\phi(y)$。

在数据完整性检测中，利用同态标签的同态性进行数据完整性验证。同态标签的性质有：

（1）使用少量特定的数据块进行数据的完整性检测，而无须对所有的数据块进行验证，从而有效降低了计算开销和通信开销；

（2）对任意两个数据块m_i和m_j，m_i+m_j的信息标签$T(m_i+m_j)=T(m_i)+T(m_j)$。

2．开源Hadoop平台

Hadoop是一种分布式系统基础架构的开源平台，用户可以在不清楚分布式的底层细节下，也用于开发分布式程序，再利用集群进行高速运算与存储。Hadoop包含两部分，分别是分布式文件系统HDFS和MapReduce实现。

HDFS采用Master/Slave的体系结构，由一个NameNode节点和多个DateNode节点组成，NameNode节点是主控服务器，主要执行对文件系统的空间命名操作，并决定数据块到DateNode节点的映射，DateNode节点主要是存储数据，并执行NameNode节点的命令。

MapReduce主要分为Map函数和Reduce函数，Map函数主要工作是把任务分解成多个任务，而Reduce函数则是把分解后的任务处理结果进一步进行汇总，得到最后的结果。

3．密钥参数生成

密钥参数生成算法是生成一对强素数，生成私钥强素数p和q，再计算公钥N和e，其中$N=pq$，生成完成后将数据返回给云用户。

智能配电网中数据的安全存储架构

智能配电网控制与数据采集系统（SCADA）中的数据主要分为实时数据和历史数据两种。HBase数据库可以实现实时数据的存储和实时数据查询，并构成分布式实时数据库系统。Hive数据仓库负责实现对海量历史数据的存储管理和历史数据查询。

配电网中数据完整性验证方案

1．配电网中数据中心模型建立

模型中的三大主体分别是：VS，验证服务器，主要作用是对电力数据中心云存储中的数据进行随机抽样的数据完整性验证工作，保证电力数据仓库中数据的完整性；CSP，云服务平台，主要是为电力数据中心提供云计算和云存储服务，其特点是存储空间大，计算资源丰富；第三方验证，其主要是帮助用户进行数据完整性验证，在本文验证方案中，假设第三方是一个可靠的机构。方案中，假设VS是信赖和独立的，其进行的完整性验证是正确且有保证的。则智能配电网中云存储数据完整性模型的结构图如图1所示。

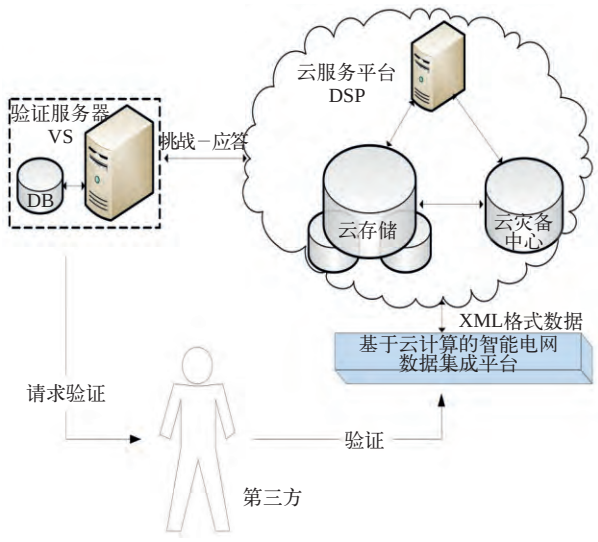

图1　智能配电网中云存储数据完整性模型

2．配电网云存储数据完整性验证方案

智能配电网数据完整性检测方案主要分为三个阶段，分别为生成密钥参数、文件预处理服务算法、数据完整性检测算法。

云用户需要上传文件时，首先调用密钥生成算法生成密钥参数，其中，私钥自己保存，公钥公开。用户将文件上传到云服务器后，云服务器对文件进行预处理操作。将文件F划分为n个

子块，为m_i（$1\leqslant i\leqslant n$），再将每个子块划分为K个基本块，为$m_{i,j}$（$1\leqslant i\leqslant n$，$1\leqslant j\leqslant k$），调用标签生成算法对每个基本块$m_{i,j}$生成标签$T_{i,j}$，文件子块$T_i=\sum_{j=1}^{k}T_{i,j}$。

（1）生成秘钥参数参数。首先，生成两个长度相同但不相等的大素数p和q，并计算$N=pq$，$\phi(N)=(p-1)(q-1)$，并生成小奇数e:e∈Z_n，且有gcd（e，$\phi(N)$），其中Z_n代表的是整数集。最后，私钥$sk=(p,q)$，公钥$pk=(e,N)$。

（2）文件预处理服务阶段。云用户上传文件F，分块参数n；分块后的文件块为m_i（$1\leqslant i\leqslant n$），计算文件块的标签信息为$T_i=e^{m_i}modN$，并将文件块和相应的标签信息存储。

（3）数据完整性检测服务阶段。用户向第三方审计发送要检测完整性的文件F，选取文件F包含的文件块m_i（$1\leqslant i\leqslant n$），并生成随机数r，计算$e_r=e^r modN$。

云服务器根据发送的挑战信息，计算检测信息1：$R=(e_r)^{\sum_{i=1}^{n}m_i}modN$。计算$S=(\pi_{i=1}^{n}T_i)modN$，并将R和S作为应答发送给第三方。第三方收到应答信息后，计算检测信息2：$R'=S'modN$。最后计算R和R'是否相等，若相等则返回云用户TRUE，否则返回FLASE。

算例分析

1．正确性分析

为验证本文算法的正确性，我们对其正确性进行证明。

证明：由$T_i=e^{m_i}modN$，得

$$S=(\pi_{i=1}^{n}T_i)\ modN=\pi_{i=1}^{n}m_i modN=R$$

则：

$$R'=S^r mod$$

$$e^{r\sum_{i=1}^{n}m_i}modN$$

$$(e_r)^{\sum_{i=1}^{n}m_i}modN=R$$

由此得出结论，本文的云存储数据完整性验证算法是正确的。其中e为小奇数，i为数据块的索引数；S为多个T_i的参数。

2．安全性分析

由于本文算法的安全性是基于大整数因式分解的困难性。即在大整数因式分解困难性问题的假设下，本文数据完整性检测方案是安全的。

假设攻击者能对模N进行分解为p和q，那么攻击者就能轻易地破坏本文算法，但是大整数因式分解是十分困难的，通过随机的选择两个512位的素数并求积，在可行的时间内是不可能实现的。因此，本文算法具有安全性。

3．数据动态性

本方案支持数据动态更新，即云用户在任何时间都可以对存储在云端的数据进行数据增加、更新和删除操作，并且可靠的第三方可对云端中数据的完整性进行验证。

假设用户要修改存储在云端的数据m_i（$1\leqslant i\leqslant n$）为m_i。首先，用户上传需要修改文件的索引号和相应的位置。然后云服务器对修改后的文件m'_i计算标签信息$T'_i=e^{m'_i}modN$。最后，云服务更新数据块m'_i和标签信息T'_i。如果用户想确认是否更新完成，通过向第三方发送一个验证请求来进行确认，由于任何的数据都会产生一个新的标签，所以更新后的数据块一定会发现是否进行改动。

假设用户要插入一个数据m'_i。用户首先需要将新插入的数据上传到云服务器，然后云服务器对新上传的数据进行签名，得到标签信息$T'_i=e^{m'_i}modN$。最后，云服务器将数据库m'_i和标签信息T'_i更新到云服务器中。云用户若要验证存储在云端的数据是否正确，发送一个验证信息给第三方后，第三方便可以对云端数据进行数据完整性验证。

假设用户要删除云存储中的数据m'_i。用户首先向云服务器发送删除文件的索引号和相应的位置。云服务收到删除请求后，将索引号对应的数据删除。云用户为了验证云服务器是否正确删除了数据m'_i，向第三方发出验证请求进行数据完整性验证，第三方将验证的结果发送给云用户。

结束语

在智能电网时代，收集反映电网运行和发展进程的大数据，并对这些数据进行实时处理是十分有价值的。例如通过有效分析智能仪表采集到的大数据，有效地帮助决策部门获得电力用户的分类特征和行为特征等信息，有助于更好的解决电力用户的需求等问题。为了保证智能配电网中数据安全可靠的存储，本文提出一种面向智能配电网的数据完整性验证方法，使用同态标签和Hadoop相结合，通过正确性和安全性分析证明本文方法的有效性。在本文的基础之上，应进一步研究如何提高验证效率的问题，这将是下一步的主要研究内容。EA

（收稿日期：2015.02.27）

（编辑：白文亭）