基于大数据分析的用户信息多重加密存储技术

陈贵平[1] 王子牛[2]
(贵州师范大学大数据与计算机科学学院 贵阳 550001)[1]
(贵州大学大数据与信息工程学院 贵阳 550025)[2]

摘 要 在大数据的分析中,当前方法对信息进行加密存储时,主要以线性微分求解对混合加密存储方法进行优化。在对密钥进行扩展的过程中,信息链路加密存储的信息出现非线性的突变,造成加密存储的信息安全性较低。鉴于此,提出一种基于超带宽的用户信息多重加密存储方法。利用混沌映射给用户信息增加反破解的保护外壳,以达到大数据分析下用户信息多重加密的目的,克服了当前方法存在的弊端,降低了加密信息存储产生的非线性突变。利用超带宽多重加密存储技术对用户信息进行多重加密存储,有效地增强了加密存储信息的抗攻击性,提高了用户信息多重加密存储的安全性,完成了对基于大数据分析的用户信息多重加密存储技术的研究。实验结果表明,利用该方法进行信息的多重加密存储提高了信息的安全性。

关键词 大数据分析,用户信息,加密存储技术,混沌参数调制映射

中图法分类号 TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.07.025

Multiple Encrypted Storage Technology of User Information Based on Big Data Analysis

CHEN Gui-ping[1] WANG Zi-niu[2]
(School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550001, China)[1]
(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)[2]

Abstract In the analysis of big data, when information is encrypted and stored, the method of linear differential solution is used to optimize the mixed encrypted storage method. In the process of the key expansion, information encrypted and stored in information link has a nonlinear mutation, resulting in low security of the encrypted storage. This paper proposed a multiple encrypted storage method for user information based on ultra bandwidth. Chaos mapping is used to increase the protection shell of user information to achieve the purpose of multiple encryption of user information under big data analysis. The disadvantages of current methods are overcomed, and the nonlinear mutation of encrypted information storage is reduced. Multiple encrypted storage is implemented for user information by using super bandwidth multiple encryption storage technology, which effectively enhances the anti-attack of the encrypted storage information, improves the security of the multiple encrypted storage of the user information, and completes the research on multiple encrypted storage technology of the user information based on big data analysis. The experimental results show that the security of information is improved by using this method to store multiple encrypted information.

Keywords Big data analysis, User information, Encrypted storage technology, Chaos parameter modulation mapping

1 引言

用户信息存储技术作为大数据分析下信息技术的核心内容,推动着IT业中技术的共同发展,是当今IT领域中的热点[1-2]。信息加密是保护用户隐私与保证信息机密性的唯一方式。其方法是对大数据分析的信息进行加密,并将解密的密钥发送至信息共享的用户。数据信息的拥有者不能为在线的其他用户提供密钥,而只有两种选择:1)依赖信任的第三方;2)将密钥分发给服务的提供商[3]。依赖信任的第三方会增加隐私问题,需要确保第三方不会将用户的隐私用于其他目的[4-5]。分发密钥也存在隐私被侵犯的风险,对此,可对基础公钥机制进行重新设置,使数据信息的拥有者与使用数据信息者的公钥对密钥加密,并由信任的第三方将加密的密钥分发给合法的用户[6]。但这种方法在扩张性方面存在严重的问题,当信息共享的用户增加时,数据信息的拥有者需要进行非对称的加密工作[7]。

如果信息代表生命与生存,那么信息的价值是不可估量的,而存储作为数据信息的载体,会使信息价值得以增值与实

到稿日期:2017-10-09 返修日期:2018-01-03 本文受2016年贵州省科学技术协会研究项目:基于大数据技术的黔归人才现状分析与研究(201602号),贵州省教育厅高等学校人文社会科学研究项目:贵州高校师范类毕业生教师专业发展状况研究(13GH052号)资助。

陈贵平(1979-),男,硕士,副教授,主要研究方向为教育信息技术、大数据、信息安全,E-mail:jy203c@foxmail.com(通信作者);王子牛(1961-),男,硕士,副教授,主要研究方向为信息处理与分析。

现，存储的数据信息是现代社会真正的财富，如何对其进行安全、有效的保存是当前需要研究的重大课题[8]。

杜朝晖等[9]提出了一种属性加密技术的安全数据存储方案。该方案利用索引的表达能力来确保数据的安全性；通过服务提供商运行时的开销与其他用户参与信息数据的检索服务，使加密运算能代替搜索，同时能使搜索的过程与数据库的管理兼容。实验结果表明，与其他方法相比，该方法搜索信息的速度较快，具有较好的性能，能保护用户的信息隐私，但安全性较差。文献[10]提出了属性加密的中心控制模型。其基于 CP-ABE 分发，设计了多授权的中心属性的方案，提高了数据密钥的安全性；设计了最小化的属性分组算法，使用户访问数据时能根据需要进行密钥的分配，减少了对不必要的属性密钥分配，且减少了数据加密属性的数量，同时增加了读写的属性，加强了加密对文件访问的控制。安全性方面的分析仿真表明，用户访问请求的响应时间较短，但安全性较差。

2　基于大数据分析的用户信息多重加密存储技术研究

利用混沌的映射给用户的信息增加反破解的保护外壳，实现了大数据分析下用户信息的多重加密，克服了当前方法存在的弊端，降低了加密信息存储产生的非线性突变；利用超带宽多重加密存储技术对用户信息进行多重加密存储，有效地增强了加密存储信息的抗攻击性，提高了用户信息多重加密存储的安全性，完成了对基于大数据分析的用户信息多重加密存储技术的研究。具体步骤如下。

2.1　用户信息多重加密

用户信息的多重加密可分为用户信息发送端的加密与信息接收端的加密，信息发送端是由映射相结合而形成的，信息接收端则由二进制编码序列对相关加密参数进行调制，利用一级生成状态的序列状态值与参数进行调制相互结合形成的。

二进制的数据信息 s_n 调制在映射参数上：

$$y_{n+1}=\lambda(s_n)y_n(1-y_n),\lambda(s_n)=\begin{cases}\lambda_a, & s_n=1\\ \lambda_b, & s_n=1\end{cases} \tag{1}$$

其中，y_n 表示映射的状态；$\lambda(s_n)$ 表示进行调制的数据信息参数，$3.59<\lambda(s_n)<4$；$4\delta=\lambda_a-\lambda_b$ 表示高低的电平差，用户信息的多重加密的密钥也由此提供，设为 (y_0,λ_b,δ)。

一级信息映射则是通过对二级映射调制进行变换调制所获得的。二级映射的用户信息参数表示为：

$$\mu^*=\frac{0.41[y_n-0.0625\lambda_a^2(4-\lambda_a)]}{0.25\lambda_a-0.0625\lambda_a^2(4-\lambda_a)}+3.59 \tag{2}$$

用户信息映射的加长导致的迭代区间表示为：$(0.0625\lambda_a^2(4-\lambda_a),0.25\lambda_a)$。

二级的信息映射状态值在调制的作用下表示为：

$$x_n^*=|0.65\cdot(x_n-0.9y_n)|\bmod(0.25\mu^*) \tag{3}$$

其中，x_n^* 表示受控的二级信息映射的状态。式(3)是在置乱的基础上考虑用户信息的状态值得出的。

从得出的信息状态值的序列能看出，二级映射在一级映射调制的作用下不具备用户的混沌性，但可表示为：

$$x_{n+1}=[\mu^*x_n^*(1-x_n^*)+kv_n]\bmod(0.25\mu^*) \tag{4}$$

由式(4)迭代得出用户信息的密文序列为：

$$\{e_n|e_n=\mu^*x_n^*(1-x_n^*)+kv_n|\}$$

其中，μ^*，x_n^* 由式(2)与式(3)得出。式(4)中存在伪消息 v_n，将其与消息的密钥一起发送给用户，除了可增加映射的迭代复杂性外，还可转移攻击者的目标，从而使欺骗性有所增强。

接收信息用户通过密钥解密，解密的核心是信息发送的映射，触发型受控对一级映射参数进行信息调制。二级映射信息产生的序列值对接收的序列值进行对比判决。

假定用户信息数据受到“污染”，该方法依照聚类的原则进行判决机制的建立。

定义的信息度量的举例表示为：

$$d(e_n,\hat{e}_n)=W((e_n)_2\oplus(\hat{e}_n)_2) \tag{5}$$

其中，$W()$ 表示求取的信息序列值的二进制中编码序列的个数，e_n 表示接收的信息序列值，$\hat{e}_n$ 表示产生的信息同步的序列值。通过分析数据信息，如 $(e_{k1},\hat{e}_{k0})_{20}>5$，其中 e_{k1} 与 $\hat{e}_{k0}$ 表示算法序列中二进制不同的信息序列值，度量的信息尺码长度为 20 位。

信息判决的步骤如下：

1)在进行第 k 次迭代的用户判决中，如果 $d(e_k,\hat{e}_n)<5$，则可进行步骤 3)；如果 $d(e_k,\hat{e}_n)\geqslant5$，则可以触发受控的电平，使电平产生一次跳变，在进行 $k-1$ 次迭代运算后，在迭代结果 $\hat{e}_{k-1}$ 的基础上进行第 k 次迭代运算并转至步骤 2)。

2)利用新求出的 e_k 再次计算信息的距离，如果 $d(e_k,\hat{e}_{k-1})\geqslant5$，能判断为故障，反之转步骤 3)。

3)如果 $d(e_k,\hat{e}_{k-1})\leqslant2$，以 $\hat{e}_k$ 为基础准则，依据对应的信息调制确定接收的比特 $\hat{s}_k$，维持 $k+1$ 次迭代的信息调制结果可返回步骤 1)，否则进行步骤 4)。

4)如果 $2<d(e_k,\hat{e}_k)\leqslant4$，则暂停运算，向用户信息发送端发送第 k 次信息迭代序列值。

2.2　基于超带宽的用户信息加密存储技术

基于超带宽的用户信息加密存储技术由超带宽的存储及梯度质量的控制构成。

1)根据用户对存储数据信息使用强度 $T(x)$ 与信息带宽的占用概率 $P(x)$ 进行超带宽的信息存储控制。

由于用户信息在大数据分析中是独立多重加密的，不同的信息节点数据的强度是不同的，因此可由信息使用的强度 $T(x)$ 与信息带宽的占用概率 $P(x)$ 进行传输与匹配。独立的信息存储的节点 t_0 进行实时的信息接收，信息数据流由指数 λ 决定，信息使用的强度 $T(x)$ 可满足指数分布。信息使用的强度 $T(x)$ 的一阶矩 $E[T(x)]$ 可满足：

$$E[T(x)]=\frac{\lambda-E|T(x)|}{\lambda^2E|T(x)^2|-E|T(x)|-e} \tag{6}$$

其中，$E|T(x)|$ 表示 $T(x)$ 的信息期望，$|T(x)^2|$ 表示 $T(x)$ 信息的二阶期望。

信息的多重加密存储的使用强度 $P(x)$ 的一阶矩阵 $E[T(x)]$ 与 $T(x)$ 的一阶矩阵 $E[T(x)]$ 呈线性的反比例关系，表示为：

$$E[P(x)]=\frac{\lambda^2E|T(x)^2|-E|T(x)|-e}{\lambda-E|T(x)|} \tag{7}$$

若由式(6)获取的一阶矩阵是正数,可说明用户信息的加密存储过程为正向分布,信息存储可通过信息链路进行带宽传输;如果获取的一阶矩阵为负数,则需要采用相应的机制,对信息的粒度进行调整,从而明显加强数据的传输强度,增加信息的安全性。

2)根据最小的信息传输的粒度 p' 进行超带宽梯度的信息传输。

当最小信息粒度所触发的信息加密过程受到条件式(6)获取的一阶矩阵的影响时,需要对信息的粒度进行调整,通过用户信息的粒度对信息存储使用的强度 $T(x)$ 进行信息带宽的控制。

$T(x)$ 与 p' 呈负相关,当最小的信息传输的粒度触发时,$T(x)$ 的正负在信息传输的周期内不会产生变化,表示为:

$$T(x)\Rightarrow p' \tag{8}$$

式(8)中,$T(x)\Rightarrow p'$ 的一阶矩阵 $E[T(x)\Rightarrow p']$ 能满足随机的莱斯分布,该分布信息的数学特征函数需满足:

$$E[T(x)\Rightarrow p']=\int E[T(x)] \tag{9}$$

假设下一时刻的信息最小传输粒度依然为 p',则信息使用的强度 $T(x)$ 可满足:

$$T(x)=\int E[T(x)]+E[T(x)\Rightarrow p'] \tag{10}$$

考虑到用户信息最小的传输粒度,信息存储的梯度 Δ 可实现信息覆盖 $T(x)$ 的全适应。对于任意时刻 Δt,用户信息梯度的系数 $\Delta T(x)$ 能满足:

$$\Delta T(x)=p'\int \sqrt{\Delta^2-T(x)^2}\,\mathrm{d}x \tag{11}$$

其中,p' 表示用户信息最小的传输粒度。

由此可得用户信息多重加密存储的指数 $\Delta\lambda$ 满足:

$$\Delta\lambda=\frac{E|T(x)|-\lambda}{1-\lambda^2E|T(x)^2|-E|T(x)|}p'\times \int \sqrt{\Delta^2-T(x)^2}\,\mathrm{d}x \tag{12}$$

用户信息梯度的弹性系数 Δ 与 $T(x)$ 需要满足下述条件,以进行二次加密存储:

$$T(x)=\Delta(1-\Delta)(\sqrt[\Delta]{1-\lambda}) \tag{13}$$

由上述过程完成基于大数据分析的用户信息多重加密存储技术的研究。

3 实验分析

在 Matlab 7 的仿真环境下进行实验,操作设备为 PC 机,搭载平台为 NetFPGA,操作系统为 contos5,内存为 4GB。

对超带宽的用户信息多重加密存储进行验证分析,可从该方法的信息占用空间、能耗以及安全性方面进行考虑。将本文方法与 Logistic 信息加密存储方法、Lorenz 信息加密存储方法进行对比与分析。图 1 给出了不同方法的用户信息多重加密存储的占用空间。对图 1 进行分析可知,本文方法 ROM 信息空间的占用节数居中,而 RAM 的信息空间占用的节数与其他方法相同。本文方法经过映射迭代运算,信息占用空间较为理想,能满足信息加密所需空间大小的要求,减少了信息对空间的占用,增加了信息存储的空间。图 2 给出了不同方法的信息加密存储的能耗。

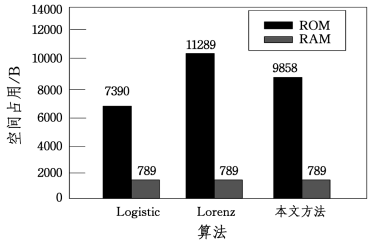

图 1 不同方法的占用空间

Fig. 1 Comparison of space occupancy of different methods

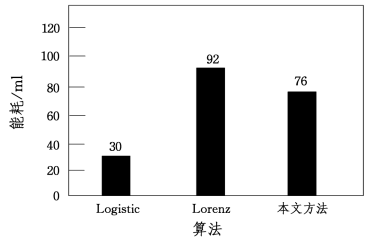

图 2 不同方法的信息加密存储能耗对比

Fig. 2 Comparison of information encryption storage energy consumption of different methods

对图 2 进行分析可知,本文方法的信息加密存储的能耗处于 Logistic 信息加密存储方法与 Lorenz 信息加密存储方法之间,由此说明,算法越简单,信息加密存储过程中的能耗就越低。本文方法为一阶映射与二级映射的结合,因此信息加密存储的能耗处于其他两种方法之间,有助于信息加密存储,提高加密信息的抗攻击性。

有效的信息加密存储方法需要注重安全性,图 3 和图 4 给出了本文加密前的用户信息字符与加密后的信息字符。

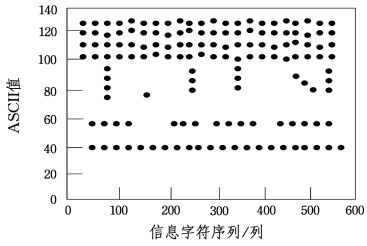

图 3 加密存储前的信息 ASCII 值的分布

Fig. 3 Distribution of information ASCII value before encrypted storage

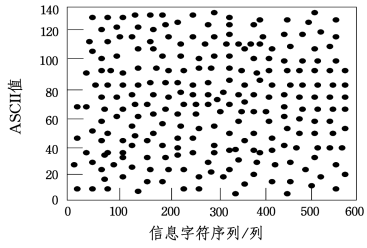

图 4 加密存储后的信息 ASCII 值的分布

Fig. 4 Distribution of information ASCII values after encrypted storage

本文方法充分利用了混沌参数调制映射的随机性与扩散

性，使加密存储后的信息特性发生了变化。使用本文方法对用户的信息进行加密存储测试。由图 3 和图 4 可以清晰地看出，未进行加密存储的信息具有明显的排列规律，而利用本文方法对加密信息进行存储，信息序列被打乱，表示为混沌的随机状态，ASCII 值的分布较为均匀，掩盖了原始的信息，有效地抵制了攻击，提高了信息多重加密存储的安全性。

结束语　为解决大数据分析中用户信息的加密存储技术存在的安全性较差的问题，提出了基于超带宽的用户信息加密存储方法。实验表明，利用该方法能有效地降低信息加密存储的复杂性，提高信息的安全性。下一步，将提高各信息节点在存储时的性能，降低信息加密存储使用的强度，增强本文方法的使用价值。

参 考 文 献

[1] SONG K，WU H J. Cloud storage based on privacy protection and efficient micro-encryption scheme[J]. Application of Electronic Technique，2016，42(7)：111-113. (in Chinese)
宋可，吴宏建. 云存储中基于隐私保护的高效的微型加密方案[J]. 电子技术应用，2016，42(7)：111-113.

[2] LEI L，CAI Q W，JING J W，et al. Enforcing Access Controls on Encrypted Cloud Storage with Policy Hiding[J]. Journal of Software，2016，27(6)：1432-1450. (in Chinese)
雷蕾，蔡权伟，荆继武，等. 支持策略隐藏的加密云存储访问控制机制[J]. 软件学报，2016，27(6)：1432-1450.

[3] LI J，LI J F，FANG F. Research of File Encryption Storage and Deletion Mechanism in Cloud Storage[J]. Journal of Chinese Computer Systems，2015，36(4)：836-839. (in Chinese)
李杰，李景峰，房方. 云存储中文件加密存储和删除方法研究[J]. 小型微型计算机系统，2015，36(4)：836-839.

[4] PAN Q H. A Double Thread Complementary Information Encryption Algorithm Based on Random Amplitude Modulation [J]. Bulletin of Science and Technology，2015，31(12)：144-146. (in Chinese)
潘期辉. 采用随机码幅度调制的双线程互补信息加密算法[J]. 科技通报，2015，31(12)：144-146.

[5] WANG S，LU Y，CHEN L Y. Research of mixed encryption algorithm in cloud storage[J]. Electronic Design Engineering，2016，24(23)：54-57. (in Chinese)
王双，卢昱，陈立云. 云存储中的混合加密算法研究[J]. 电子设计工程，2016，24(23)：54-57.

[6] CHENG X X，HAN X Z，CHEN X J，et al. An Optimization Encryption Algorithm for Cloud Storage and Its Simulation[J]. Computer Simulation，2016，33(4)：356-359. (in Chinese)
程肖肖，韩宪忠，陈雪蛟，等. 一种用于云存储的优化加密算法的研究与仿真[J]. 计算机仿真，2016，33(4)：356-359.

[7] LU Y，WANG S，CHEN L Y. Research of Mixed Encryption Algorithm Based on Cloud Storage[J]. Computer Measurement & Control，2016，24(3)：129-132. (in Chinese)
卢昱，王双，陈立云. 基于云存储的混合加密算法研究[J]. 计算机测量与控制，2016，24(3)：129-132.

[8] HANG T X，DING J Y. A Cloud-storage Privilege Revoking Optimizing Mechanism Based on Dynamic Re-encryption[J]. Science Technology and Engineering，2015，15(20)：108-115. (in Chinese)
韩同欣，丁建元. 基于动态重加密的云计算存储平台权限撤销优化机制[J]. 科学技术与工程，2015，15(20)：108-115.

[9] DU Z H，ZHU W Y. Implementation of secure data retrieval schema in cloud storage by using ABE technology[J]. Application Research of Computers，2016，33(3)：860-865. (in Chinese)
杜朝晖，朱文耀. 云存储中利用属性基加密技术的安全数据检索方案[J]. 计算机应用研究，2016，33(3)：860-865.

[10] QIU S W，LI Y Y. Reliable data delivery with low delay in energy harvesting wireless sensor network[J]. Journal of Computer Applications，2015，35(2)：345-350. (in Chinese)
邱树伟，李琰琰. 能量捕获无线传感器网络中低时延的可靠数据传递[J]. 计算机应用，2015，35(2)：345-350.

[11] HU X D，CAI D Q. Design and research of secure encryption clustering algorithm for wireless sensor networks[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)，2009，21(3)：421-424. (in Chinese)
胡向东，蔡东强. 无线传感器网络安全加密成簇算法的设计及研究[J]. 重庆邮电大学学报(自然科学版)，2009，21(3)：421-424.

(上接第 138 页)

[10] QIAN Q，XIAO C J，ZHANG R. Propagation modeling for P2P botnet in structured P2P network [J]. Journal of Software，2012，23(12)：3161-3174. (in Chinese)
钱权，萧超杰，张瑞. 结构化对等网络中 P2P 僵尸网络传播模型[J]. 软件学报，2012，23(12)：3161-3174.

[11] OUYANG C X，TAN L. New propagation model of Botnet on scale-free network [J]. Computer Engineering and Applications，2013，49(9)：110-114. (in Chinese)
欧阳晨星，谭良. 无尺度网络下的僵尸网络传播模型研究[J]. 计算机工程与应用，2013，49(9)：110-114.

[12] CAO X L，NIU Z L. Study on propagation model of botnet based on weighted networks [J]. Computer Applications and Software，2012，30(7)：180-184. (in Chinese)
曹晓丽，牛志玲. 基于加权网络的僵尸网络传播模型研究[J]. 计算机应用与软件，2013，30(7)：180-184.

[13] CHENG S P，TAN L，HUANG B，et al. Botnet propagation modeling and analysis [J]. Computer Engineering and Applications，2013，49(1)：107-111. (in Chinese)
成淑萍，谭良，黄彪，等. 僵尸网络传播模型分析[J]. 计算机工程与应用，2013，49(1)：107-111.

[14] SRICHARAN K G，KISORE N R. Mathematical model to study propagation of computer worm in a network[C]// 2015 IEEE International Advance Computing Conference (IACC). IEEE，2015：772-777.

[15] REN W，SONG L P，FENG L P. A novel mathematical model on Peer-to-Peer botnet [J]. Journal of Measurement Science and Instrumentation，2014，5(4)：62-67.

[16] BUONO C，VAZQUEZ F，MACRI P A，et al. Slow epidemic extinction in populations with heterogeneous infection rates [J]. Physical Review E，2013，88(2)：022813.