文章编号：2096-1472(2018)-06-44-03
DOI:10.19644/j.cnki.issn2096-1472.2018.06.013

基于大数据分析的网络安全系统优化算法改进研究

于淑香[1]，王 浩[1,2]
(1.沙洲职业工学院电子信息工程系，江苏 张家港 215600；
2.苏州大学计算机科学与技术学院，江苏 苏州 215006)

摘 要：大数据分析环境下的网络安全系统优化算法应用，具有严格的控制原则。本文主要针对大数据环境下，网络安全数据检测算法应用形式进行探讨，并针对大数据分析过程中容易产生的网络安全隐患问题进行总结。可以作为网络安全系统优化算法构建应用的技术参照，同时本文所提出的程序汇编控制方法具有实践应用意义。经过大数据网络安全系统优化算法改进实验，最终得到一种局域网络环境下更安全高效的运算处理系统。

关键词：大数据分析；网络安全；系统优化
中图分类号：TP301 **文献标识码**：A

Improvement Research on the Optimization Algorithm of the Network Security System Based on Big Data Analysis

YU Shuxiang[1],WANG Hao[1,2]
(1.*Department of Electronic and Information Engineering*,*Shazhou Professional Institute of Technology*,*Zhangjiagang* 215600,*China*;
2.*School of Computer Science and Technology*,*Soochow University*,*Suzhou* 215006,*China*)

Abstract: In the environment of big data analysis,the application of network security system optimization algorithm is conducted in accordance with strict control principles.In this paper,the application form of network security data detection algorithm is discussed in the big data analysis environment,and the potential network security risks in big data analysis are summarized.The research results can provide technical reference for the application of the network security system optimization algorithm,and the program assembly control method proposed has practical application significance.Through the experiments on the optimization algorithm of the network security system based on big data analysis,the study finally achieves a more secure and efficient computing and processing system in the local area network environment.

Keywords:big data analysis;network security;system optimization

1 引言(Introduction)

大数据分析环境下对网络系统进行优化构建，最常使用的安全检测算法为模式匹配法、机器学习算法和指纹检测算法。大数据分析环境下，算法指标提取是实现网络安全优化的重要环节，对于算法指标提取控制，需要从大数据环境中确定一个分析范围，所选取的算法指标应该具有针对性。能够与网络安全系统中的各项参数相匹配，保持一致共同实现大数据分析[1]。数据片段选择可以作为网络安全系统优化算法开展的针对对象，数据片段选择不仅要从高效运算角度出发，更应该考虑处于大数据网络环境中，数据自身存在的威胁。随着数据分析下载任务进行发现风险隐患也能及时停止，避免将带有病毒的数据下载到网络资源存储设备中，数据库对接还需要选择高效算法为连接基础，算法帮助下数据资源之间存在差异性也能利用极端时间来调整一致。数据库在网络环境下，可以帮助提升算法应用安全性，数据库对接和安全防护作用下所开展的各项算法应用也将顺利进行。网络风险与数据库传输中的安全控制选择，应该体现出数据库的存储能力，并对数据存储中的各项资源，优先选择安全部分来计算。

2 大数据分析环境下的网络安全数据流检测算法(Network security data flow detection algorithm in big data analysis environment)

(1)模式匹配算法

模式匹配算法根据大数据网络环境下所得到的数据流，与网络系统内预先设定的数据库进行匹配，匹配结果保持一致的数据可以进入到优化运算中。模式匹配运算方法，并不是针对单个字符来进行的，而是能够将字符整理成数据流集合。用于处理复杂的数据分析，通过数据匹配可以节省大量时间，模式匹配属于深度开发算法，会在匹配过程中表达出不同数字符号特征，并将大数据分析过程中所传输的参数，整理成为特征一致的字符，这样在传输和共享阶段便不会出现匹配误差。网络环境中存在的安全威胁形势多变，因此在对算法进行选择时，应该体现出灵活多样性，可以根据不同使用方向自动调整，整理成为安全的字符控制模式。

模式匹配算法中对于字符表达采用DFA模式，该种方法属于一种正表达模式，会根据数据产生顺序来进行匹配。该

基金项目：江苏省教育厅高等职业教育产教深度融合实训平台建设项目(编号：201618).

种方法具有较快的匹配速度，在运算效率上可以满足大数据环境下网络安全需求，但对系统内存要求十分高，需要同一DFA集合中，能够同时满足不同种数据分析变化。

(2)机器学习算法

机器学习算法属于一种自动更新模式，能够根据网络环境下所遇到的不同隐患，来自动匹配算法防护模式。机器学习算法需要安全口令来进行验证，通过安全验证的信息数据更新，下载需求才可以得到满足。机器学习法最常用在DPI系统上，数据流安全检测中，可以在网络环境中自动获取数据更新包，并按照安全协议来进行下载资源与数据库之间的匹配。匹配成功后视为所下载数据与系统安全内容相同，接下来的数据自动更新计划也将能够继续完成。当机器学习算法应用中检测发现数据流内存在风险隐患，会对此类信息特征进行采集，并整理成为集合保存在数据库中，针对风险集合所构建的数据库。在网络信息环境下会自动发出警告，同时对已经识别的安全数据流进行防护，避免网络病毒入侵到机器学习算法核心控制系统内[2]。该算法在数据流分类能力上是其他模式所难以达到的，由于需要对认定的风险项目进行二次识别，因此在机器学习算法应用中，运算消耗时间也比较大，适合应用在机械生产自动化控制中。

(3)指纹检测算法

指纹检测算法属于网络安全系统中的一种验证模式，最常用在移动网络设备中，可以对用户启动网络系统身份进行识别与验证。当以用户身份发出请求时，在指纹识别检测结果中表现出异常。该种算法会从两个，方面来进行安全检测，首先是对指纹验证通过的用户，会将所检测得到的数据流自动匹配到安全访问模块中。另一部分没能通过指纹验证的用户请求，将会被自动隔离，视为风险项目作出处理。大数据环境下网络安全系统优化算法的改进，主要是针对运算速度提升与安全防护精准度来开展。为了可以不用在end+1−min_location>=w时，遍历w窗口中的值，采用一个2*2的二维数组存储当前的最小值、最小值后面的次小值及其位置。需要注意的是，此算法需要考虑这样一个问题——当cur与min_location之间的距离大于w时，需要从次小值的位置到cur位置间重新选取最小值和次小值，而且要保证次小值在最小值的后面。大数据处理技术正在逐渐向移动网络设备层面开发应用，不仅要对常规互联网进行安全防护，更应该考虑移动网络设备特征，形成一种符合移动网络设备使用中安全防护需求特征的算法。指纹识别对于数据流的匹配过程比较简单，会将预先存储的指纹特征与请求指纹特征进行对比，相似度达到安全认证标准后便能够通过请求。指纹库生成如图1所示。

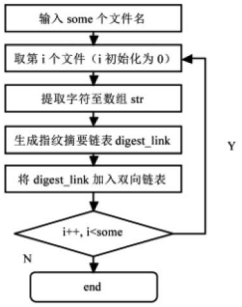

图1 指纹库生成流程

Fig.1 Fingerprint database generation process

3 大数据分析环境下网络安全系统优化算法改进 (Optimization algorithm of the network security system in the big data analysis environment)

(1)算法指标提取

为确保大数据分析速度，大数据分析中可以预先截获数据流，针对于网络安全系统中数据库内信息相匹配的资源优先运算，并确定出文本边框界定范围生成数据库后，利用网络信息传输环境，来构建文本框[3]。针对网络环境下，可能会产生的风险隐患，形成多样性防护系统。假设文本块字数为n，首先使用Karp−Rabin算法公式，生成n−k+1个数字。对n−k+1个数，选取前W个数字(1，···，W)中的靠右边的最小值，下一次选取第2个到第W+1个数字中的最小值，依次进行下去，一直到第n−w+1个字符。将这些最小值设为边界，得到(b1，b2，b3，b4，…)。回到原文本块中，取文本块中的第b1个字符到第b2个字符间的文本块。然后通过MD5计算文本块的Hash值，即指纹。优先选择大数据环境下的可靠信息片段，利用文本框算法边界确定缩小运算时的峰值范围，大数据环境下每一项运算任务进行所针对的数据，都应该保障运算顺序，以免在数据类型划分中出现干扰情况。确定数据分析运算顺序后，接下来进行的数据归类也能高效进行，并不会产生数据之间的干扰。数据库文本框的应用是对数据资源使用中，运算内容的确定，文本框确定作用下，不会产生数据重复运算的问题。

(2)数据片段选择

通过数据片段优化选择，体现出系统内部更为高效的控制内容，数据片段选择，可以结合大数据分析中文本框界定来进行。综合探讨出不同方向对于大数据系统选择的需求，并通过文本框优化构建出符合大数据环境功能实现的网络安全系统。由于Winnow算法每次都是从下一个字符寻找长度为W的窗口中的最小值，造成很大的计算冗余。当取得第1个数到第w个数间的最小值min时，min值前面的数字肯定比min值大，也就是说，在后面的min次比较中，min值前面的数肯定不是最小值，那么只需要比较min值和其后的数字即可。但是，为了保证获得完整的边界值，在第二次比较时，应该比较起始点为min，终点为w+1间的数。由此可得，假设当前窗口min值的位置是location，窗口的末端位置为end；下一次仅比较min_location的值同end+1的值即可；针对移动网络设备所开展的数据优化，在对数据片段确定时，同样需要体现不同数据构建过程中对于网络风险隐患的控制能力。综合大数据分析处理需求，对原有算法体系作出优化，尤其是移动网络设备，更应该考虑网络系统构建自身对数据算法的需求。对移动网络与互联网两个不同方向的算法选择进行综合控制，移动通讯设备自身存储空间比较小，如果算法应用中占用大量空间将会影响最终的数据传输，对此数据片段尽最有效部分进行应用，避免传输速度受到影响[4]。

(3)数据库对接

数据库对接是实现网络资源下载的基础，将网络设备中的数据库与实时数据传输系统进行结合，利用大数据环境下信息处理速度提升特征，为网络信息选择和数据优化处理提供环境。有关于数据对接传输中可能会产生的风险隐患，在算法优化时也会重点处理，仅仅依靠预防控制方法是很难

达到最佳效果的。因此数据优化还应该从实质性内容层面开展，尤其是针对大数据分析运算中的网络安全需求，算法优化时为避免出现故障，可以采用实时对接的模式进行运算。安全检测系统数据库对接如图2所示。

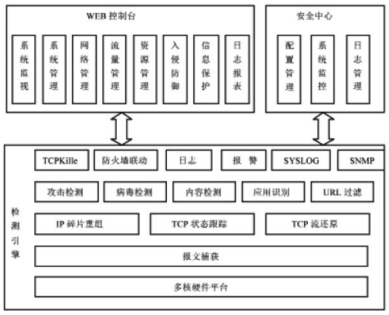

图2 基于DPI安全检测系统架构图

Fig.2 DPI security detection system architecture diagram

4 大数据分析环境下网络安全系统优化算法的实现(Implementation of the optimization algorithm of the network security system in the big data analysis environment)

(1)数据存储

大数据分析环境下安全系统优化算法应用后，需要对功能实现情况进行检验，首先从数据存储层面来进行。判断在数据存储过程中，是否能够对网络环境中存在的风险进行有效规避，对于数据存储和传输期间可能会产生的功能隐患，构建安全防护基本框架便可以得到解决[5]。在安全防护基本框架中所存在的各项功能隐患问题，会通过系统文本框形式展现出来。针对大数据分析网络安全所开展的数据运算任务，也能在文本框规定范围内精准实现，对于文本框与大数据环境之间的对接融合，提升数据存储能力可以对安全运算对接匹配进行检验。一旦数据存储内容与网络环境中下载的资源存在误差，通过安全优化算法，也可以更高效地控制解决。大数据分析技术应用图如图3所示。

大数据分析环境下网络信息传输，需要体现出移动网络与互联网络两种不同形式，目前wifi无线网络也是一种网络接入点形式，在系统中更应该考虑这一特征，并体现出综合控制对网络信息传输中的误差补偿效果。

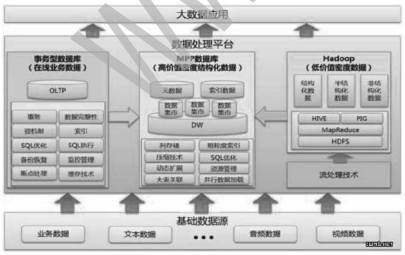

图3 大数据分析技术应用处理图

Fig.3 Big data analysis technology application processing diagram

(2)文本框划分

文本框划分同样建立在软件程序汇编基础上，软件程序汇编会确定详细的范围，通过文本框划分来避免数据运算中相互干扰影响进度的情况。文本框中所存储的信息是有网络环境中实时下载更新的，基于互联网络环境下开展的数据信息存储和各项网络资源更新，利用无线网络传输。可以实现同一端口连接多个设备的使用目标，同时对于数据流的生成，以及连接应用，在数据库更新防护作用下，也可以达到最佳安全控制效果。对于网络传输与数据实时更新控制，文本数据库在内容形式上始终保持自动学习更新状态，在网络环境下更高效地下载并使用资源，数据运算处理任务实现不会受到影响。对于网络环境下的各项资源检测，可以利用自主学习技术来优化数据资源更新效果。网络数据库资源使用，学习能力是首先要保障的，通过学习能力引入，来实现大数据分析环境下网络安全系统设法控件的创新，无论是文本框使用还是最终的数据资源库建立，在学习技术帮助下都能够确定明确的方向目标。

(3)数据流生成

数据流生成可以对网络安全系统功能进行强化，利用网络环境中的资源优化更新，观察数据流生成后，与预期的功能效果之间是否存在误差。生成不同安全性检测曲线，并利用数学建模来对曲线内的参数节点进行优先运算，建立起与文本框相匹配的数据资源库。数据流生成后行成与网络系统安全需求相匹配的资源，建立在移动网络设备自动安全检测信息技术上，数据流能够自动完成控制内容匹配，充分建立起安全检测文本框架，实现大数据分析环境下不同信息安全检测需求的实时匹配。发现匹配结果与分析内容之间产生隐患，也能够在大数据控制基础上，进行规划整合。当所下载得到的数据存在大幅度误差和波动时，也能够通过数据文本框更新来进一步控制。数据流划分并得到验证后，网络系统安全性已经完成初步判断，对于一些比较容易出现的潜藏隐患，在系统内部会有针对性的构建并用，利用大数据环境中综合控制处理技术，来全面提升数据流的更新下载安全性。

5 结论(Conclusion)

虽然本文提出的对指纹算法的改进可以检测出在网络上传输的数据是否含有敏感信息泄露问题，但是不能解决检测具有相同语义的敏感信息是否也被泄露。当文本发生较大规模的更改，或者只是将字面做了更改，语义没有变化的时候，就不能检测出敏感信息的泄露问题了。该方面内容仍然需要继续强化研究。

参考文献(References)

[1] 何欢.群智能算法优化神经网络在网络安全的应用[J].现代电子技术,2016,39(20):12-14.

[2] 齐银峰,谭荣建.基于改进粒子群优化算法的BP神经网络在大坝变形分析中的应用[J].水利水电技术,2017,48(2):118-124.

[3] 王延纬,刘凡,丁涛,等.含多分区电力系统能量备用实时优化调度的动态积极集求解算法[J].西安交通大学学报,2017,51(4):45-52.

[4] 刘洋,刘博,王峰.基于ParameterServer框架的大数据挖掘优化算法[J].山东大学学报(工学版),2017,47(4):1-6.

[5] 叶娟.基于网络数据包的大数据分析程序设计与开发[J].软件工程,2018,34(1):35-37.

作者简介：

于淑香(1978-)，女，硕士，讲师.研究领域：软件工程.

王 浩(1977-)，男，博士，讲师.研究领域：人工智能，软件工程.