

[配电网]

# 一种面向配电网无线公共信道接入的安全检测算法

文/广东电网有限责任公司梅州供电局 李舒 范群滔  
华北电力大学 刘俊杰

针对配电网无线公网存在的安全缺陷,为基于无线公网的配用电业务承载网络提供高安全、高可靠和易扩展的安全防护体系、认证机制及安全防护策略。提出一种融合家族优生学的K-means聚类算法,结合了K-means聚类算法突出的分类能力的特点与家族优生学最优选择的优点,提升了配电网无线公网安全检测的性能,为配电网的安全运行创造了有利的条件。

随着智能配电网的发展,终端规模不断扩大,无线公网技术在很大程度上弥补了有线通信技术的不足。在一些不具备铺设光纤条件或光纤铺设不经济的地区,可采用无线公网技术作为补充,保证接入网的高覆盖率和投资的经济性。

目前无线公网在配电网的应用越来越普及也越来越多样化,这使得配电网网络数据不仅数量庞大而且数据结构非常复杂。为了保证无线公网网络数据的安全性,网络攻击对电力行业所造成的危害极为严重;不仅可能导致电力运营数据的泄露,间接导致国家生产经济数据的泄露以及用户身份信息的泄露,甚至造成生产经营中关键系统、系统配置信息、生产经营数据信息的删除或损坏,势必产生巨大的社会负面影响。因此,配电网公网需要对网络中的数据进行更好的安全检测极为必要,通过配用电业务及智能设备接入方式及无线公网通道监测和安全防护等方面实现技术创新,切实提高配电网通信网的安全运行水平。

数据量大且结构复杂意味着配电网无线公网具有复杂性、动态性等特点,其网络完全性具有一定的脆弱性,来自内外部的安全风险给网络安全工作带来了极大的压力与挑战。而随着计算机技术与网络业务的发展,配电通信网络安全将面

临着新业务与网络威胁的双重挑战,也对配电网无线公网的安全检测算法提出了更高的要求。

针对此情况,本文提出一种融合家族优生学的K-means检测算法,可以有效提高大数据环境下集中选取数据准确性的能力,实现网络安全检测准确率与检测效率的提升,从而为提升配电网通信网的安全运行水平创造有利条件。

## 安全检测攻击算法的选择

在当前配电网无线公网网络数据量巨大且结构复杂的环境下,要对大量不同类型和结构的数据进行准确、高效、快速的挖掘是一件非常困难的事,由此,对于网络安全检测和网络安全防护来讲具有非常大的压力。对数据类型中具有代表性的数据进行分析,采用有效的数据检测算法,来提高网络检测的准确性就具有非常重要的意义。聚类算法是一种以群分析的数据挖掘算法,其将数据集划分为若干模式子集,并将具有较高相似度的数据集中在同一集合中,不同集合之间具有较为明显的属性差异性。

配用电无线公网网络环境认知的攻击检测的主要实现方法是在网络运行过程中节点随环境的变化随时根据需要插入新的攻击检测代码,目前

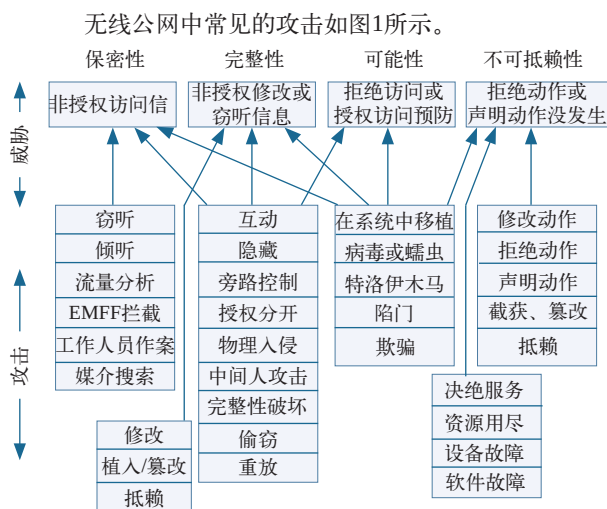


图1 常见网络攻击分类

1967年，由J.B.Mac Queen提出的K-means算法是聚类算法思想中的一种。K-means算法是根据数据的层次将数据分类划分，每一类数据都具有极高的相似度，以其相似度的平均值为标准将其划分为 $k$ 个聚类。在网络安全检测中，所检测的数据不仅数量多，而且结构复杂。要从网络数据 $n$ 个对象中随机选择 $k$ 个对象作为初始聚类中心，并要求每一个对象的属性都具有较为鲜明的特点，彼此之间存在较大的属性差异化，将所有数据对象按照与初始聚类中心的距离划分聚类，以这些对象到它们的距离为划分标准，分别将它们分配给与其最为相似的聚类；其次，通过计算获取新的聚类，并对新聚类所有对象的中心求均值，此计算过程反复的重复，不断的形成新的距离，再求均值，直到均方差所标记的标准测度函数开始出现收敛为止，以达到最佳准确度的均值。

K-means算法是一种基于样本间相似性的聚类算法。此算法以 $k$ 为参数，将在网络大数据中所随机提取的 $n$ 个对象分为 $k$ 个簇，这些簇的特点是具有较高的属性相似度，并且簇之间的属性具有明显的不同，以此为方法对每一个簇中的对象求平均值，已达到将其归为某一准确类。通过K-means聚类算法将配电网无线公网中传输的数据进行聚簇分类后，可以将系统中的合法信息进行分类，将入侵信息筛选出来，从而达到保护配电网中传输数据的完整性和安全性的目的。

在传统的K-Means算法中，通常采用均方误差作为标准测量函数

$$J = \sum_{i=1}^K \sum_{x_j \in C_i} |x_j - m_i|^2 \quad (1)$$

式（1）中 $x_j$ 是空间的一个数据点； $C_i$ 为聚类； $m_i$ 为聚类的均值。

虽然k-means聚类算法有很多的优点，但是从其聚类的过程来看此方法的缺点也很明显，如 $k$ 值需要在聚类前给定，对初始中心点的依赖性很大，对于较大数据量，算法的时间开销非常大等。可以说这些缺点在某种程度上限制了k-means聚类算法在实际中的应用。

当聚类数据对象为密集型数据，并且所有数据对象各个类之间具有较明显区别时，采用传统K-means进行聚类效果较好，能满足网络安全检测需求。但是在实际应用中，入侵检测网络数据包通常都为随机实时抓取的，因此要事先确定出聚类数，进行初始聚类中心选取则较为困难，而 $k$ 值不准确，则聚类结果不稳定。另一方面，传统的K-means算法只能处理连续型数值，不能处理离散型数值，应用范围狭窄。

为了提高K-means聚类算法的分类能力，选取最优的聚类中心，我们将家族优生学与K-means聚类算法进行融合。利用K-means聚类算法突出的分类能力的特点，结合家族优生学最优选择的优点，提升配电网无线公网安全检测的性能，为配电网的安全运行创造了有利的条件。

### 融合家族优生学的K-means聚类算法

为了提升配电网网络安全检测能力，本文提出的融合了家族优生学的K-means检测算法，采用家族优生学算法选取最优聚类中心点，通过K-means聚类算法对配电网无线公网数据进行检测。融合家族优生学的K-means聚类算法步骤如下：

1) 产生初始种群:随机生成由 $m$ 个可行解组成的初始数据群体。

对于配电网无线公网中的数据，通过家族优生学选取出 $K$ 个最优的聚类中心点 $C_i$ ，作为 $k$ 个簇初始质心。

2) 计算适应值:计算初始数据群体中每一个体的适应值。

将所有配电网无线公网中传输的数据指派到最相似的簇，根据K-means聚类算法，采用平方误差准则计算与 $k$ 个初始质心的距离，平方误差定义如式（2）所示

$$E = \sum_{i=1}^K \sum_{p \in C_i} (p - C_i)^2 \quad (2)$$

其中,  $p$  是配电网无线公网传输的数据,  $c_i$  是簇  $C_i$  的质心。当平方误差达到最小时, 聚簇结果满足下面的条件

$$\forall p \in PC, \text{distance}(p, \text{getCluster}(p)) \leq \min_{0 < i \leq m} \{\text{distance}(p, C_i)\} \quad (3)$$

其中,  $p$  表示配电网无线公网传输的数据样本,  $PC$  表示配电网无线公网传输的数据集合,  $\text{distance}(p, C_i)$  表示配电网无线公网样本数据与聚簇中心的距离,  $\text{getCluster}(p)$  表示样本所属聚簇中心,  $M$  表示聚簇个数,  $C_i$  表示第  $i$  个聚簇。

3) 建立配对池。基于适应值建立群体数据中每一个体的选择概率, 依此概率分布, 对群体作  $M$  次随机选取, 将选出的个体放入“配对池”。

4) 组成家庭。配对池中个体两两结合形成  $M/2$  个家庭(每一家庭中高适应值个体为父本, 另一个为母本)。

5) 双亲优化交配。对每一家庭, 双亲以正交优化交配方式产生若干“中间子女”。

如表1所示给出一种简单的正交表  $L_4(2^3)$ , 安排3元素2水平实验。表1中A、B、C代表三个因素, 数字“1”“2”代表因素的两个不同水平。这种安排从  $2^3$  种可能组合中选4种组合进行实验。

表1 正交表  $L_4(2^3)$

| 试验次数 | A | B | C |
|------|---|---|---|
| 1    | 1 | 1 | 1 |
| 2    | 1 | 2 | 2 |
| 3    | 2 | 1 | 2 |
| 4    | 2 | 2 | 1 |

通过正交表可以用较少的试验得到较好或最好的试验结果。

6) 子女竞争。通过家庭内部竞争, 仅2个子女成长为下一代的合法成员。

7) 变异。由合法子女组成的群体接受变异操作。

8) 聚类。根据选取出的最优中心值进行K-means聚簇分类。

再次采用平方误差准则计算与  $K$  个优化后质心的距离为

$$E' = \sum_{i=1}^{K'} \sum_{p \in C_i} (p - C'_i)^2 \quad (4)$$

式中,  $E'$  为重新计算的平方误差,  $K'$  为经过家族优生学选取后的簇心个数,  $C'_i$  为经过家族优生学选取后的簇心集合。根据  $E'$  的值对配电网无线公网中的数据进行检测。

算法流程图如图2所示。

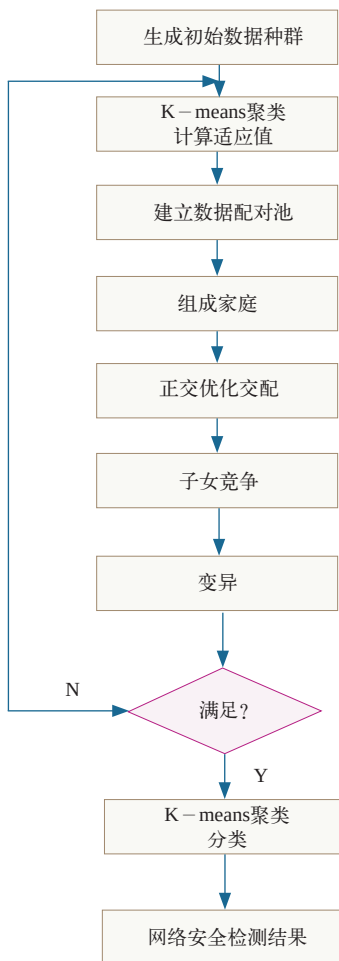


图2 融合家族优生学的K-means聚类算法流程图

通过家族优生学的算法选出K-means聚类算法的最优聚类中心  $C_i$ , 再通过K-means聚类检测算法对配电网无线公网中的数据进行检测。有效提高对配电网无线公网中传输数据的分类能力, 从而进一步提升系统检测入侵信息的能力。

## 结束语

本文针对当前配电网无线公网网络数据量大且结构复杂, 其接入方式安全性不高的环境下, 提出了融合家族优生学的K-means聚类算法, 采用家族优生学算法选择聚类中心, 对传统的K-means聚类算法实现了优化。从而有效提升了配电网无线公网的安全检测能力及配电网系统的安全防御能力, 使配电网无线公网信息的安全性和完整性得到了保证。EA

(收稿日期: 2018.07.08)

(编辑: 白文亭)