doi:10.11835/j.issn.1000-582X.2020.249

面向 IEC61850 智能变电站的网络安全异常流量分析方法

王 胜[1],唐 超[1],张凌浩[1],张 颉[1],王 海[1],柴继文[1],刘珊梅[1],郑永康[1],
邓 平[2],曹 亮[3],夏晓峰[4],秦 帆[4]

(1. 国家电网四川省电力科学研究院,成都 610072;2. 国家电网自贡供电公司,四川 自贡 643000;
3. 国家电网甘孜供电公司,四川 甘孜 626700;4. 重庆大学 大数据与软件学院,重庆 400044)

摘要:为了保证智能变电站的网络通信安全和整个变电站的稳定运行,提出了一种基于机器学习 k-means 聚类算法的异常流量分析方法。根据智能变电站中过程层网络的特性,结合对 IEC61850 智能变电站专有 GOOSE(generic object-oriented substation event)以及 SV(sample value)协议的报文结构解析,使用了一种基于信息熵的特征选取方法对智能变电站正常工作时站内网络通信流量进行特征分析选择,利用 k-means 聚类算法完成了对异常流量的检测分析及其相关分析。相较于以往方法,文中方法对智能变电站的过程层网络流量信息的特征进行了选取,根据信息熵理论,完成了重要特征的选择和冗余特征的剔除,提高了聚类算法的效率,提高了对异常流量检测的准确性。

关键词:智能变电站;IEC61850;k-means 聚类;异常流量

中图分类号:U448.213 **文献标志码**:A **文章编号**:1000-582X(2022)01-001-08

Research on network security abnormal flow analysis method for IEC61850 intelligent substation

WANG Sheng[1], *TANG Chao*[1], *ZHANG Linghao*[1], *ZHANG Jie*[1], *WANG Hai*[1], *CHAI Jiwen*[1], *LIU Shanmei*[1], *ZHENG Yongkang*[1], *DENG Ping*[2], *CAO Liang*[3], *XIA Xiaofeng*[4], *QIN Fan*[4]

(1. Department of Information and Communication Security and Technology, Sichuan Electric Power Research Institute, Chengdu 610072, P. R. China; 2. State Grid Zigong Power Supply Company, Zigong 643000, Sichuan, P. R. China; 3. State Grid Ganzi Power Supply Company, Ganzi 626700, Sichuan, P. R. China; 4. School of Big Data & Software Engineering, Chongqing University, Chongqing 400044, P. R. China)

Abstract: In order to ensure the network communication security of intelligent substations and their stable operation, this paper proposes an analysis method of abnormal flow based on machine learning k-means clustering algorithm. Firstly, according to the characteristics of the process level network in the intelligent

收稿日期:2020-02-23

基金项目:国网四川省电力公司科技资助项目(52199717001P);国网四川省电力公司电力科学研究院资助项目(SGSCDK00XTJS1800093)。

Supported by Science and Technology Project of State Grid Sichuan Electric Power Company(52199717001P) and Project of Electric Power Research Institute of State Grid Sichuan Electric Power Company (SGSCDK00XTJS1800093).

作者简介:王胜(1987—),男,工程师,主要从事信息安全研究,(E-mail)240517810@qq.com。

substation, the message structure of IEC61850 intelligent substation's proprietary GOOSE(generic object-oriented substation event) and SV protocol is analyzed. Then, the network communication flow in the intelligent substation during normal operation is analyzed and selected by using a feature selection method based on information entropy. Finally, k-means clustering algorithm is used to complete the detection and analysis of the abnormal flow. Compared with the previous methods, the proposed method first selects the characteristics of process layer network flow information of intelligent substation. According to the theory of information entropy, the selection of important features and the elimination of redundant features are then completed, improving the efficiency of clustering algorithm and the accuracy of abnormal flow detection.

Keywords: intelligent substation; IEC61850; k-means clustering algorithm; abnormal traffi

随着计算机技术和网络技术的发展,特别是互联网及社会公共网络平台的快速发展,在"两化"融合的行业发展需求下,为了提高生产和管理效率,电力行业大力推进工业控制系统自身的集成化、集中化管理,智能电网也就随之而来。变电站作为电力系统即智能电网中的重要组成部分,担负着电压转换、电能分配、输配电的控制和管理等重要任务,其安全、可靠地运行是整个智能电网安全和稳定的重要保障。

智能变电站由于其自身的重要性,关系到某一片区所有的电力供应。一旦智能变电站某一功能模块出现细微差错,所造成的损失将是不可估量的。如今也有许多不法分子对智能变电站进行恶意攻击,使智能变电站出现故障,导致整个电网进入瘫痪状态,造成巨大的经济损失,2019年委内瑞拉大停电事件就是一个很好的例子。

智能变电站的安全是多方面的,除了上述的外部恶意攻击以外,由于变电站设备高度的集成化和一体化,内部工作人员的一些不当操作、站内设备的突发意外故障等也都会导致智能变电站的安全面临严峻的挑战。

如今智能变电站通常采用IEC61850标准[1],其内部所有变电站的信息共享与数据传输通过智能变电站的过程层网络来实现。过程层网络的报文是变电站内部各功能部件间相互协调的通信载体[2],它包括了整个变电站所运行的状态以及相应的设备信息。因此,过程层网络的稳定与正确性是整个智能变电站得以正常运行的重要保证。这也是文中所研究的重点所在。

笔者介绍了智能变电站网络的结构特性,对智能变电站安全进行了详细分析,阐述了国内外的研究现状,提出了面向智能变电站过程层网络安全异常流量检测的方法。

1 相关工作

1.1 智能变电站网络

在智能变电站中,根据IEC61850标准主要采用的是"三层两网"的架构体系[3],如图1所示。

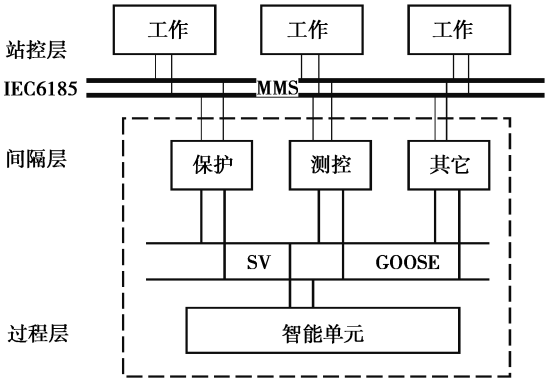

图1 智能变电站"三层两网"结构

Fig. 1 Three layers and two networks of intelligent substation structure

“三层”分别是“站控层”“间隔层”和“过程层”，“两网”则就是“站控层网络”和“过程层网络”。其中，“站控层网络”主要基于 TCP/IP 的 MMS 协议。通过 MMS 协议，实现站控层对间隔层的信息统一管控。“过程层网络”则是直接连接间隔层和最底层的过程层设备，是直接管理控制过程层设备运行的。它主要使用专有的工控传输协议“GOOSE”和“SV”。

过程层的设备主要包括合并单元、智能终端，以及变压器和发电机等，并且直接面向电力系统的设备，所以过程层网络传输的稳定性和正确性，是保证整个变电站和电力系统可以稳定运行的关键。文中工作也是针对过程层网络的流量进行分析检测。

1.2　过程层网络

在过程层网络中，智能变电站主要使用报文来进行信息的传输。报文内容的正确性和传输的实时性，是变电站正常运行的重要保障。一旦过程层网络报文流量传输出现异常，将会直接导致过程层所面对的电力系统的一次设备出现故障或者差错。所以对过程层的流量进行检测和分析十分重要。过程层网络所传输的主要是 GOOSE 和 SV。GOOSE 报文主要承载的是面向对象的变电站事件信息，而 SV 报文主要承载的信息则是采样值的信息。

1.2.1　GOOSE

GOOSE 是工业控制系统中的专有协议，它是面向通用对象的变电站事件，不需要经过回执确认，由系统直接选择顺序重发机制，具有实时性和一发多收的特点。GOOSE 保存的信息主要是一些保护设备的信息，如跳闸、闭锁等[4]；以及变电站的时间信息。GOOSE 报文的发送和接收分别通过 publisher(公告式发布)和 subscriber(预订式接受)执行。每当变电站中有设备数据发生变化，GOOSE 都会即刻进行传送。并且在此阶段，GOOSE 的发送间隔会逐渐地增大，直至数据的变化趋于稳定即事件状态稳定。GOOSE 的报文帧格式如表 1 所示。

表 1　GOOSE 报文格式

Table 1　Message structure of GOOSE

字段	内容说明
HeadMAC	MAC 目的地址(6 字节)
	MAC 源地址(6 字节)
Prioritytagged	TPID(2 字节标记)＝0x8100
	TCI(2 字节)
网络数据类型	Ethertype(2 字节)＝0x88B8
	APPID(2 字节)＝0x0000～0x3FFF
	Length(2 字节)＝8＋m(m 是 APDU 长度)
	Reserved1(2 字节)＝0x0000
	Reserved2(2 字节)＝0x0000
	APDU

1.2.2　SV

SV 所承载的是设备的采样值信息，以网组传播订阅的方式传输。SV 会周期性地将设备的电流电压的采样信号进行传输。一旦有数据丢失，会立即重新传送。SV 在网络上传输时采用的是开放式系统互联通信参考模型(open system interconnection reference model)，即 OSI 模型，但是只用到了其中的 4 层，包括应用层、表示层、数据链路层和物理层。传输层和网络层为空，在应用层将协议数据单元 PDU 定义好，经过表示层的编码后直接映射到数据链路层和物理层，这样避免了通信堆栈而造成的传输延时，保证了报文传输和处理的快速性。SV 的报文格式如表 2 所示。

表2 SV报文格式

Table 2 Message structure of SV

字段	内容说明
HeadMAC	MAC目的地址(6字节)
	MAC源地址(6字节)
Prioritytagged	TPID(2字节标记)=0x8100
	TCI(2字节)=0x0000
网络数据类型	Ethertype(2字节)=0x88BA
	APPID(2字节)=0x4000～0x7FFF
	Length(2字节)=8+m(m是APDU长度)
	Reserved1(2字节)=0x0000
	Reserved2(2字节)=0x0000
	APDU

1.3 安全分析

从智能变电站的本身来看,智能变电站中的一些二次设备十分脆弱。根据IEC61850标准所建立的智能变电站,其内部设备的通信安全缺少有效的保护手段。尤其是GOOSE和SV协议,在设计时对信息安全的考虑是不够完善的,GOOSE和SV采用多播的形式进行传输,针对这一点,至少存在着DoS和重放等攻击手段可以对智能变电站内部的通信进行有效地攻击。一旦智能变电站受到了上述任何一种攻击,整个智能变电站的网络环境就会受到巨大的影响,根据1.1节中对智能变电站的过程层网络的概要介绍,无论是GOOSE还是SV协议,对网络传输的实时性都有严格的要求。

在智能变电站中很难对业务流量进行分析,对潜在的异常流量行为模式无法做到完全的感知。因此,很难防范一些异常行为的出现,一旦这类异常行为出现,可能会对二次设备造成很大的损坏,所造成的损失难以估计[5]。

所以,对智能变电站的过程层流量及异常流量进行检测,对整个智能变电站的安稳运行有着重大意义。

1.4 研究现状

在中国,随着电力需求的发展以及国家对电网发展战略的制定,全国各地已经陆续开始了许多智能变电站的建设。从2010年6月发生的“震网”病毒等事件后[6],学术界意识到电网安全的重要性以及相应预防措施的缺乏性,并且也提出了一系列的解决和缓解方案。比如,杨雅辉等[7]提出了一种基于增量式的GHSOM神经网络模型的入侵检测研究方法;Yu等[8]提出通过分析过程层设备异常以及它们所引起的信息流,也包括信息流异常状态的形成机理和相应特征量,提出并设计了不同的异常保护判据;丁修玲等[9]也提出了一种基于报文解析的变电站过程层网络异常信息流的保护方法,通过电力系统和通信系统的联合仿真证实了电网中通信延迟的危害。也有许多研究直接从工业控制系统出发,通过对系统的权限、认证等角度入手,从而保证了智能变电站的安全性[10]。

在已有的相关技术方法中,对异常流量的检测,通常是从时间、空间上构造“正常模式”或“正常流量”作为基准,或者是通过设置简单阈值来判断异常流量。前者并没有结合报文内的信息,而后者由于变电站网络中会存在突发报文,阈值的设置很难选择[11]。文中采用了文献[12]中的特征选取方法,完成对网络流量报文内与流量类型相关的重要特征的选择,剔除冗余特征,使分类效果得到提升。文中提出对智能变电站仿真平台进行攻击实验,从而对所收集的正常运行流量和突发的异常流量通过k-means聚类算法进行分类,可以有效地识别过程层网络中的异常流量。有监督的机器学习方法可以建立一个准确的模型来进行检测,由于它需要对大量的样本进行人工标记,成本过大、工作量过多,所以研究选用典型的非监督的机器学习算法k-means聚类算法。

2　面向智能变电站过程层网络安全的异常流量检测

2.1　过程层网络异常流量特征选择

异常流量特征的选取主要使用了文献[12]中信息熵理论的特征选取法，不仅可以得到流量数据中的重要特征，而且也能够去除冗余特征，使检测结果更加准确。

2.1.1　引入信息熵度量特征

“熵”是热力学中的相关概念。在 1948 年，香农首次把热力学中有关“熵”的概念引入到信息论之中，从而把“熵”作为了一个随机事件发生的不确定性的度量。

当一个事件有可能会出现 n 种可能的结果，每种结果独立出现的概率都为 p_i 时，信息论中香农使用了度量 I 表示事件的不确定性结果

$$I=\sum_{i=0}^{n} p_i \ln p_i , \tag{1}$$

式中，事件的可能结果越多，不确定性越大，完全确定时，不确定性为零。并且，香农将与 I 成正比的

$$S=K\ln n , \tag{2}$$

称为信息熵。也就是说，信息熵越大的时候，其事件的可能性就越多，所蕴含的信息量就更大，反之亦然。

文中通过引入信息熵度量特征的重要程度。一个属性特征的信息熵越大，所包含的信息量就越大，就越能作为区分的判定。反之，如果一个属性的信息熵值很小，表明所有的流量数据在这个属性上的差异就不大，从而也就无法通过此项属性值来对流量进行区分判定。

2.1.2　流量特征选取方法

文献[12]中的特征选取方法，将特征主要分为了两类：重要特征和不重要特征。与流量类别有着相关关系的特征一般是重要特征，即该特征被引入后会对类别产生影响。而不重要特征则与此相反，表示与其不相关。通常使用 1 s 内正常流量和异常流量特征熵值的差值来区分重要特征和不重要特征，特征重要系数 K 为

$$K_A=|\text{正常流量}\overline{A_S}-\text{异常流量}\overline{A_S}| , \tag{3}$$

式中：A_S 表示某源 IP 地址的特征 A 在 1 s 内的熵值；$\overline{A_S}$ 则为平均值。若 $K_A>0.1$，则 A 为重要特征，反之为不重要特征。有关 K_A 阈值的选择参照文献[12]中的定义 2。

同时，对于冗余特征，定义了冗余系数 T，来表示 2 个特征的冗余程度：

$$T_{AB}=\min\left(\left|\frac{\overline{A_S}-\overline{B_S}}{\overline{A_S}+\overline{B_S}}\right| , |\overline{A_S}-\overline{B_S}|\right) , \tag{4}$$

式中，T_{AB} 越接近 1，则 2 个特征越不相关，反之就越相关。

如果 2 个特征互为冗余特征，只需要删除熵值平均值较小的那个特征即可。

2.2　智能变电站实时流量数据聚类分析方法

k-means 聚类算法是最为经典的使用最为广泛的一种聚类方法。它是基于距离的一种聚类算法，使用样本间的距离作为划分样本的依据，距离越近则样本就越接近，反之同理。k-means 算法流程如图 2 所示。

k-means 算法的第一步就是随机地选择 k 个初始聚类中心。k 值过大或者过小都会导致效果不佳。k 值过小，每个簇内的数据差异很可能会过大；而 k 值过大，又会使各个簇间的数据差异很小[13]。对于这个问题，遵循 Rezaee 等[14]提出的最佳 k 值的范围是$(1, \sqrt{n})$。

在上述的最佳区间内随机选择一个 k 值，并选出 k 个初始聚类中心：$Z=(Z_1, Z_2, \cdots, Z_k)$，计算各样本与各聚类中心的距离，并将其与距离最小的中心划分为一类，式(5)中 d 表示样本与样本中心的距离，s 表示所选出的特征项。

$$d=\sqrt{\sum_{j=1}^{k}(s_j-Z_{ij})^2} 。 \tag{5}$$

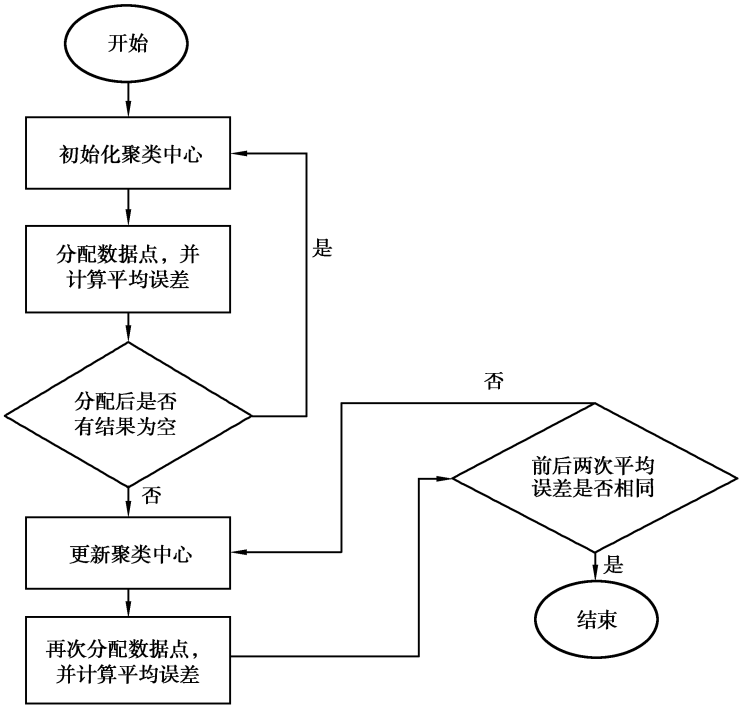

图 2 k-means 算法流程

Fig. 2 Algorithm steps of k-menas

更新聚类中心，计算出各个类簇的样本距离的平均值：

$$Z'_{ij}=\sum_{s\in Z_i}\frac{s_j}{|Z_i|}。\tag{6}$$

计算整个数据集的平方误差 S_{SSE} 判断是否收敛：

$$S_{SSE}=\sum_{i=1}^{k}\sum\nolimits_{s\in Z_i}|d(s,Z_i)|^2。\tag{7}$$

虽然上述是在一个较优区间内选择的 k 值，但是仍有可能不是最理想的 k 值，所以引入轮廓系数[15]，来选择一个最优的初始 k 值。轮廓系数是评判聚类效果好坏的一个指标，其区间为[−1,1]，其值越大效果越好。在上述区间内，随机选择 N 个 k 值，并分别计算每次的轮廓系数，选择最终轮廓系数最大的一个。个体轮廓系数 S_i 公式为

$$S_i=\frac{b(i)-a(i)}{\max\{a(i),b(i)\}},\tag{8}$$

式中：$a(i)$ 表示和 i 同类中所有数据对象之间的平均距离；$b(i)$ 表示和 i 不同类中所有数据对象之间的最小平均距离。

总的轮廓系数为各个体轮廓系数的平均值

$$S_{total}=\frac{1}{N}\sum_{i=1}^{n}S_i。\tag{9}$$

2.3 智能变电站过程层网络异常流量检测流程

研究目的是对智能变电站中的过程层网络中异常流量进行分析，从而检测到电站在运行时可能产生的异常流量，及时做出反应避免损失。k-means 算法是属于非监督型的机器学习算法，无需事先进行大量的学习工作。

对异常流量数据进行检测的过程如图 3 所示，分别是对流量数据进行收集，包括正常流量数据和经过人为干涉后所产生的异常流量数据。对收集到的流量利用 2.1.2 的特征选择的方法找到主要的重要特征，并将特征进行标准化处理，并利用聚类算法对流量进行分类，从而达到识别检测异常流量的效果。

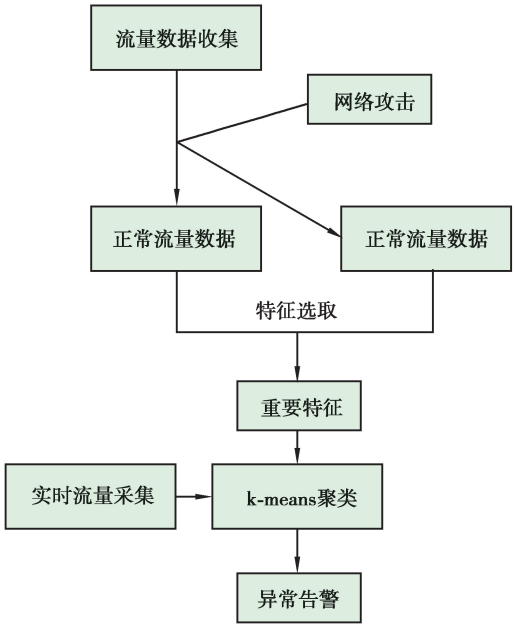

图 3　检测流程

Fig. 3　Work steps

2.4　实验验证

笔者对智能变电站的仿真平台进行抓包，对流量数据进行采集，并且在这过程中对过程层网络进行 DoS 攻击，以保证可以产生一定的异常流量。抓取一个时间段内的 GOOSE 和 SV 协议报文，将每一个 GOOSE 或者 SV 协议的报文抽象成一个网络连接，提取出 GOOSE 和 SV 报文中的相应信息，如源 MAC 地址、目的 MAC 地址、协议类型等。变电站中的所有设备的 MAC 地址和 IP 地址都是固定的，因此也可以将这 2 个信息做绑定，这样就将协议报文给抽象成了方便处理的连接数据。

在运用聚类算法对信息流量数据进行分类检测前，使用在 3.1 节中所提到的异常流量特征选择的方法来对所抓取到的流量信息进行重要特征的提取，作为聚类算法的度量标准，从而使聚类算法变得更加高效。因为所进行的是 DoS 网络攻击，所以所提取出来的最主要特征是 GOOSE 和 SV 报文中的 MAC 地址和 IP 地址以及发送间隔。

在完成了这些准备工作之后，继续采集变电站流量数据，并同时对其进行 DoS 攻击。将所获取的流量数据进行聚类检测，并对得到的聚类结果进行分析，异常簇中报文内容 92.6%与模拟的攻击报文信息吻合。笔者同时使用传统的基于阈值的方法进行了检测，结果对比如表 3 所示。

表 3　实验对比

Table 3　Experimental comparison results

方法	正确率/%	误检率/%
文中方法	92.6	7.4
阈值	79.3	20.7

3　结束语

文中介绍了智能变电站的网络结构，指出了过程层网络安全的重要性。根据过程层网络通信协议特性分析了过程层网络的安全问题。提出了一种基于 k-means 聚类算法的智能变电站过程层网络安全异常流量的检测方法，并采用了一种基于信息熵的特征选取方法。经实验结果对比得出了文中方法的可行性及其有效性。接下来的工作将在文中的基础上，进一步研究提高检测率以及检测的实时性。

参考文献：

[1] 黄雅宣. 智能变电站的涵义及发展探讨[J]. 通讯世界，2017(7)：131-132.
Huang Y X. Discussion on the meaning and development of intelligent substation[J]. Telecom World，2017(7)：131-132. (in Chinese)

[2] 江南，纪陵，杨小凡. 智能变电站信息安全技术[J]. 电气自动化，2018，40(6)：48-51.
Jiang N，Ji L，Yang X F. Information security technology of smart substation[J]. Electrical Automation，2018，40(6)：48-51. (in Chinese)

[3] 胡斌，郭亚飞，杨彬，等. 智能变电站技术的现状与发展趋势研究[J]. 智慧电力，2018，46(3)：87-90.
Hu B，Guo Y F，Yang B，et al. Research on status and development trend of smart substation technology[J]. Smart Power，2018，46(3)：87-90. (in Chinese)

[4] 中国国家标准化管理委员会.信息安全技术-信息安全风险评估规范：GB/T 20984-2007[S]. 2007.
Standardization Administration of China. Information security technology-Information security risk assessment specification：GB/T 20984-2007[S]. 2007. (in Chinese)

[5] 韩宇奇，郭嘉，郭创新，等. 考虑软件失效的信息物理融合电力系统智能变电站安全风险评估[J]. 中国电机工程学报，2016，36(6)：1500-1508.
Han Y Q，Guo J，Guo C X，et al. Intelligent substation security risk assessment of cyber physical power systems incorporating software failures[J]. Proceedings of the CSEE，2016，36(6)：1500-1508. (in Chinese)

[6] Newman M E J. The structure and function of complex networks[J]. SIAM Review，2003，45(2)：167-256.

[7] 李孟超，王允平，李献伟，等. 智能变电站及技术特点分析[J]. 电力系统保护与控制，2010，38(18)：59-62,79.
Li M C，Wang Y P，Li X W，et al. Smart substation and technical characteristics analysis[J]. Power System Protection and Control，2010，38(18)：59-62,79. (in Chinese)

[8] Radziwill N M. Countdown to zero day：stuxnet and the launch of the world's first digital weapon[J]. Quality Management Journal，2018，25(2)：109-110.

[9] Tim G. Indegy finds out when industrial controls go bad (think Stuxnet)[J]. Network World (Online)，2016.

[10] 张其林，王先培，赵宇. 基于 IEC 61850 的变电站自动化系统连锁故障分析[J]. 电力系统自动化，2013，37(2)：61-66.
Zhang Q L，Wang X P，Zhao Y. Analysis on cascading failures of substation automation system based on IEC 6185[J]. Automation of Electric Power Systems，2013，37(2)：61-66. (in Chinese)

[11] Lu Z X，Meng Z W，Zhou S X. Cascading failure analysis of bulk power system using small-world network model[C]// 2004 International Conference on Probabilistic Methods Applied to Power Systems. IEEE，2004：635-640.

[12] 丁明，韩平平. 小世界电网的连锁故障传播机理分析[J]. 电力系统自动化，2007，31(18)：6-10.
Ding M，Han P P. Study of failure spreading mechanism in the small-world power grid[J]. Automation of Electric Power Systems，2007，31(18)：6-10. (in Chinese)

[13] 马秀娟，马福祥，赵海兴. 基于耦合映像格子的有向网络相继故障[J]. 计算机应用，2011，31(7)：1952-1955,1979.
Ma X J，Ma F X，Zhao H X. Cascading failure in coupled map lattices with directed network[J]. Journal of Computer Applications，2011，31(7)：1952-1955,1979. (in Chinese)

[14] 曲朝阳，杨琴，杨杰明，等. 基于贝叶斯网络的智能变电站风险关联模型[J]. 电力系统自动化，2016，40(2)：95-99.
Qu Z Y，Yang Q，Yang J M，et al. Risk associated model of smart substations based on Bayesian network[J]. Automation of Electric Power Systems，2016，40(2)：95-99. (in Chinese)

[15] 王力军，周凯，吴迪，等. 基于风险传递网络的智能变电站二次系统风险评估[J]. 电力系统保护与控制，2018，46(6)：97-105.
Wang L J，Zhou K，Wu D，et al. Risk assessment for smart substation secondary system using risk transfer network model[J]. Power System Protection and Control，2018，46(6)：97-105. (in Chinese)

（编辑 詹燕平）