基于业务逻辑的电力业务报文攻击识别方法

王海翔[1],朱朝阳[1],王 宇[2],张锐文[2],李俊娥[2],李霁远[3],应 欢[1]
(1. 中国电力科学研究院有限公司 信息通信研究所,北京 100192;
2. 武汉大学 国家网络安全学院 空天信息安全与可信计算教育部重点实验室,湖北 武汉 430072;
3. 国网浙江省电力有限公司电力科学研究院,浙江 杭州 310014)

摘要:针对电网测控终端的电力业务报文攻击极易造成电力一次设备误动,从而引发电力事故。电力业务报文攻击通常通过干扰正常业务逻辑达到攻击目的,已有攻击识别方法没有考虑业务逻辑,有效性比较差。因此,提出一种基于业务逻辑的电力业务报文攻击识别方法,该方法定义了电力业务逻辑状态链和黑白名单,将误用检测与异常检测方法相结合,基于业务逻辑黑白名单对业务的威胁度进行评估,并考虑电网时间风险与业务重要性,对威胁度进行修正,通过比较业务威胁度与安全阈值,实现对电力业务报文攻击的高效准确识别。给出了应用所提方法实现的一个攻击识别系统架构,并对实现后的系统进行了测试,验证了所提方法的有效性。

关键词:电力业务报文攻击;攻击识别方法;业务逻辑;状态链;电网测控终端
中图分类号:TM 73 **文献标志码**:A DOI:10.16081/j.epae.202007023

0 引言

随着智能电网信息空间与物理空间耦合的不断加深[1-2],近年来,由网络攻击导致的电网物理系统故障愈发常见,严重影响了电力系统的正常运行。如2015年底,攻击者通过获取变电站监控系统服务器的操作权限进行了恶意倒闸操作,导致乌克兰电网80 000个用户停电[3];2016年以色列电力供应系统受到重大网络攻击,迫使电力供应系统离线运行[4]。电网中用于一次系统或设备参数测量和控制的各类智能终端和装置(本文将其统称为测控终端)作为沟通信息系统与物理系统的桥梁,当其遭受通过篡改、伪造与重放电力业务报文实施的攻击时,将直接影响电力一次设备的正常运行,造成如断路器异常开断、定值修改等问题,从而引发电力事故[5-6]。因此,如何有效识别电网测控终端可能遭受的电力业务报文攻击成为亟待解决的问题。

目前,针对电网测控终端的网络攻击识别研究主要分为以下2类。一类研究将传统信息网络的网络攻击识别系统直接应用于电网测控终端的网络攻击识别[7-11],如文献[7]通过异常检测方法识别利用非电力业务报文流量实施的网络攻击;文献[8]提出基于自学习通信模式的未知攻击识别方法。另一类研究利用电力专有协议流量特征、规则或报文不同字段的相关性等方式[12-18]来实现电网测控终端的网络攻击识别,如文献[12]提出利用面向通用对象的变电站事件(GOOSE)报文网络流量特征来进行攻击识别;文献[13]提出基于IEC 60870-5-104协议的流量模式检查、各个字段的合法性与相关性检查的攻击识别方法;文献[14]利用文献[13]中所提方法对使用IEC 61850协议实现的网络攻击进行识别。上述研究能有效识别地址解析协议ARP(Address Resolution Protocol)欺骗、互联网控制消息协议ICMP(Internet Control Message Protocol) Flood和同步序列编号SYN(Synchronize Sequence Numbers) Flood等利用通用网络协议脆弱性实施的网络攻击,以及部分利用IEC 60870-5-104、IEC 61850等电力专有协议脆弱性实施的网络攻击,如GOOSE畸形报文攻击等,但无法有效识别电力业务报文攻击。

电力业务报文攻击是指攻击者通过篡改、伪造与重放电网测控终端传输的业务报文从而导致电力一次设备误动的攻击,这类攻击通常会改变正常的业务逻辑。因此,为了有效识别针对电网测控终端的电力业务报文攻击,本文提出了一种基于业务逻辑的电力业务报文攻击识别方法。不同于已有的网络攻击识别方法,所提方法通过对电网业务逻辑进行威胁度评估,实现了对电力业务报文攻击的有效识别,并将误用检测与异常检测方法相结合,同时考虑电网时间风险与业务重要性指标,降低了误报率。

1 基于业务逻辑的电力业务报文攻击识别方法

电网中的业务存在一定的逻辑关系且具有规律

收稿日期:2019-08-21;**修回日期**:2020-05-28
基金项目:国家电网有限公司总部科技项目(电网嵌入式终端漏洞挖掘与攻击检测关键技术研究)(52110418001K)
Project supported by the Science and Technology Project of State Grid Corporation of China(Key Technologies of Vulnerability Mining and Attack Detection for Embedded Terminal in Power Grid)(52110418001K)

性，而电力业务报文攻击通常会造成业务逻辑紊乱，因此，通过判断当前电网业务逻辑是否处于安全状态，可以有效识别针对电网测控终端的电力业务报文攻击。

1.1 方法框架

本文利用威胁度来标识电网业务逻辑的安全状态；设计了电网业务逻辑状态链来保存电网业务逻辑；为了将当前业务逻辑状态与已知的安全或不安全的状态序列进行对比，设计了业务逻辑黑白名单来保存已知安全性的状态序列。

误报率和漏报率是攻击检测系统最重要的2个指标。本文将误用检测与异常检测方法相结合，同时考虑电网时间风险与业务重要性指标，提出了一种低误报、低漏报的基于业务逻辑的电力业务报文攻击识别方法。误用检测方法能够对包含在黑白名单规则中的业务逻辑进行精准的威胁度评估，但无法有效判定未在黑白名单中的业务逻辑，将业务逻辑全部评估为安全（危险）会带来漏报率（误报率）较高的问题。因此，本文在误用检测的基础上，结合相似度匹配异常检测方法对未在黑白名单中的业务逻辑进行进一步的评估，以有效降低误报率；但考虑到一些正常业务逻辑与非正常业务逻辑的状态序列相似，通过相似度匹配仍难以进行有效区分，本文从攻击与正常操作的目的不同出发，引入电网风险指数与业务重要性指标对相似度进行修正，以更准确地评估未在黑白名单中当前状态序列的威胁度。

基于业务逻辑的电力业务报文攻击识别方法的基本思路如下：①基于黑白名单模式匹配误用检测方法对当前业务状态进行威胁度评估，若匹配白名单则威胁度评估为0，若匹配黑名单则威胁度评估为1；②考虑到黑白名单模式匹配的识别误报率由黑白名单的优劣决定，为了降低该方法所带来的较高误报率或漏报率问题，对未在黑白名单中的状态序列进行相似度匹配；③为了进一步降低误报率，将电网时间风险与业务重要性融入基于黑白名单相似度匹配的异常检测方法中，对相似度进行修正，以更准确地评估未在黑白名单中的当前状态序列的威胁度；④将业务威胁度与安全阈值进行对比，从而实现对电力业务报文攻击的识别。识别方法的总体流程如图1所示。

1.2 数据结构设计

考虑到电网中所包含的节点数量过多，对所有节点共同进行关联分析目前尚不具备实现可行性，同时不同控制块（指一个电力业务报文中控制的所有节点的集合）上的节点大多没有直接关联关系，因此，本文选用控制块作为划分业务的依据，即仅对同一控制块上的节点进行关联分析。

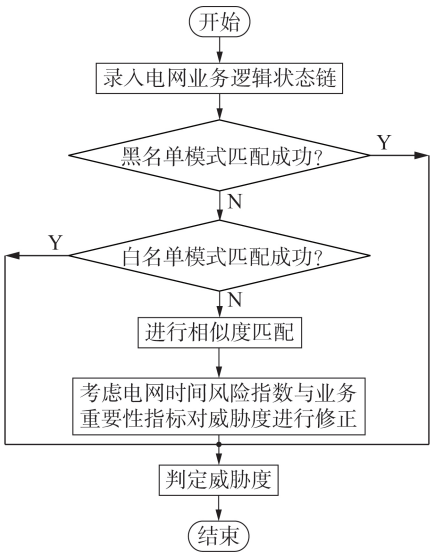

图1 所提识别方法的总体流程

Fig.1 Overall process of proposed identification method

1.2.1 电网业务逻辑状态链的数据结构

本文采用状态链来描述电网业务逻辑，包括业务状态及其变化过程，数据结构如图2所示，其由以下7个部分构成。

（1）单点信号值：即图2中的data字段，用于描述单个功能约束数据属性FCDA（Functionally Constrained Data Attribute）项的值。在电网中可以理解为刀闸开关信号或电网中某个节点的电压/电流值。

（2）信号地址：即图2中的pos字段，用于描述一个FCDA项的位置。在电网中可理解为刀闸或节点的逻辑实例名；在实际计算中可理解为一个变量的名称，用于索引该变量。

（3）多点信号地址序列：即图2中的pos_seq = $[\mathrm{pos}_1, \mathrm{pos}_2, \cdots, \mathrm{pos}_n]^{\mathrm{T}}$字段，在本文中其用于描述一个控制块上各个单点信号的信号地址序列。

（4）多点信号值序列：即图2中的status = $[\mathrm{data}_1, \mathrm{data}_2, \cdots, \mathrm{data}_n]^{\mathrm{T}}$字段，在本文中其用于描述

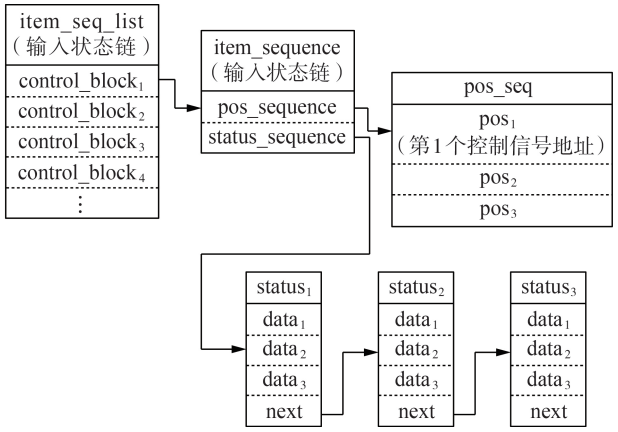

图2 状态链的数据结构

Fig.2 Data structure of state chain

一个控制块上各个单点信号值序列。在电网中可理解为多个刀闸的位置或多个节点的电压 / 电流值。

(5)状态节点：定义为 Node = {status, pos_seq}，由多点信号值序列和多点信号地址序列组成，用于描述一个控制块的状态。

(6)状态改变：用来描述一个控制块状态中的单个或多个单点信号发生变化。在电网中可以理解为一个刀闸或多个刀闸切换、一个节点或多个节点电压 / 电流值发生改变、一处或多处定值发生改变。

(7)状态序列：定义为 status_seq = {status_value, pos_seq}，其中status_value=[$status_1$,$status_2$,…,$status_n$]，由有限个 pos_seq 相同的状态节点顺序链接组成，用于描述一个控制块状态改变的逻辑过程。在电网中可以理解为一组开关的操作逻辑关系、一组电压 / 电流值发生改变的逻辑关系。

1.2.2　黑白名单的数据结构

本文采用黑白名单保存这些已知安全性的状态序列，黑白名单的数据结构见附录 A 中图 A1。黑白名单中录入以状态序列的结构表示的不同规则，其中白名单录入已知安全性的状态序列，黑名单录入已知不安全性的状态序列。

将黑名单定义为 $U_b = \{U_{b/pos\_seq_1}, U_{b/pos\_seq_2}, \cdots, U_{b/pos\_seq_n}\}$，其中 $U_{b/pos\_seq_1} = \{status\_seq_1, status\_seq_2, \cdots, status\_seq_n\}$ 为所有满足条件 pos_seq = pos_seq_1 的不合法状态序列集合。

将白名单定义为 $U_w = \{U_{w/pos\_seq_1}, U_{w/pos\_seq_2}, \cdots, U_{w/pos\_seq_n}\}$，其中 $U_{w/pos\_seq_1} = \{status\_seq_1, status\_seq_2, \cdots, status\_seq_n\}$ 为所有满足条件 pos_seq = pos_seq_1 的合法状态序列集合。

1.3　录入电网业务逻辑状态链

从电网业务报文的应用层内容中提取一个状态节点，并进行状态链录入，得到当前的状态序列 $status\_seq_{now}$，具体录入过程如下。

(1)从电网业务报文的应用层内容中提取一状态节点 $Node_{now} = \{status_{now}, pos\_seq_{now}\}$，其中 $status_{now} = [data_1, data_2, \cdots, data_k]^T$。

(2)根据状态节点 $Node_{now}$ 中的 pos_seq_{now} 找到其对应的控制块状态序列 status_seq = {status_value, pos_seq}，即满足条件 pos_seq = pos_seq_{now} 的状态序列，其中 status_value = [$status_1$, $status_2$, …, $status_{n-1}$]。

(3)判断 $status_n$ 是否等于 $status_{now}$，若相等，则录入过程结束；否则，转步骤(4)。

(4)将 $status_{now}$ 链入 status_seq 中，得到当前状态序列 $status\_seq_{now} = \{status\_value_{now}, pos\_seq_{now}\}$，其中 $status\_value_{now} = [status_1, status_2, \cdots, status_n]$。

为了简便表达，下文用 $\boldsymbol{S}'$ 表示 $status\_value_{now}$，用 $\boldsymbol{P}'$ 表示 pos_seq_{now}。

1.4　黑名单模式匹配

在状态链录入一个节点后，对录入的当前状态序列进行黑名单模式匹配，具体流程如下。

(1)根据 $\boldsymbol{P}'$ 找到对应的黑名单规则子集 $U_{b/\boldsymbol{P}'} = \{status\_seq_{b_1}, status\_seq_{b_2}, \cdots, status\_seq_{b_t}\}$。

(2)将 $status_{now}$ 添加到缓存状态 $\boldsymbol{S}''_{b/\boldsymbol{P}'} = [status_{n_1+1}, status_{n_1+2}, \cdots, status_{n_1+x}]$ 中，其中 n_1 的初始取值为 0，那么添加后 $\boldsymbol{S}''_{b/\boldsymbol{P}'} = [status_{n_1+1}, status_{n_1+2}, \cdots, status_{n_1+x}, status_{now}]$。

(3)顺序遍历黑名单规则子集中的所有规则，即对于 $\forall status\_seq_{b_i} = \{status\_value_{b_i}, pos\_seq_{b_i}\} \in U_{b/\boldsymbol{P}'}$，其中 $status\_value_{b_i} = [status_1, status_2, \cdots, status_{n_2}]$，若 $n_2 > n - n_1$ 且 $i \neq t$，则继续遍历；若 $n_2 > n - n_1$ 且 $i = t$，则匹配失败，过程结束；若 $n_2 < n - n_1$，则转步骤(4)；否则，转步骤(5)。

(4)保留 $\boldsymbol{S}''_{b/\boldsymbol{P}'}$ 中的后 n_2 项，即 $\boldsymbol{S}''_{b/\boldsymbol{P}'} = [status_{n-n_2+1}, status_{n-n_2+2}, \cdots, status_{now}]$，然后将 $n - n_2$ 置为 n_1，即 $\boldsymbol{S}''_{b/\boldsymbol{P}'} = [status_{n_1+1}, status_{n_1+2}, \cdots, status_{now}]$。

(5)判断 $\boldsymbol{S}''_{b/\boldsymbol{P}'}$ 是否与 $status\_value_{b_i}$ 相同，若相同，则只保留 $\boldsymbol{S}''_{b/\boldsymbol{P}'}$ 的最后 1 项，即 $\boldsymbol{S}''_{b/\boldsymbol{P}'} = [status_{now}]$，同时将 n_1 置为 $n-1$，返回电力业务威胁度 $P_t = 1$，匹配成功，过程结束；否则，继续遍历，即转步骤(3)。

1.5　白名单模式匹配

在黑名单模式匹配失败后，需要对当前状态序列进行白名单模式匹配，具体流程如下。

(1)根据 $\boldsymbol{P}'$ 找到对应的白名单规则子集 $U_{w/\boldsymbol{P}'} = \{status\_seq_{w_1}, status\_seq_{w_2}, \cdots, status\_seq_{w_e}\}$。

(2)将 $status_{now}$ 添加到缓存状态 $\boldsymbol{S}''_{w/\boldsymbol{P}'} = [status_{n_3+1}, status_{n_3+2}, \cdots, status_{n_3+y}]$ 中，其中 n_3 的初始取值为 0，那么添加后 $\boldsymbol{S}''_{w/\boldsymbol{P}'} = [status_{n_3+1}, status_{n_3+2}, \cdots, status_{n_3+y}, status_{now}]$。

(3)顺序遍历白名单规则子集中的所有规则，即对于 $\forall status\_seq_{w_i} = \{status\_value_{w_i}, pos\_seq_{w_i}\} \in U_{w/\boldsymbol{P}'}$，其中 $status\_value_{w_i} = [status_1, status_2, \cdots, status_{n_4}]$，若 $n_4 > n - n_3$ 且 $i \neq e$，则转步骤(7)；若 $n_4 > n - n_3$ 且 $i = e$，则匹配失败，过程结束；若 $n_4 < n - n_3$，则转步骤(4)；否则，转步骤(5)。

(4)保留 $\boldsymbol{S}''_{w/\boldsymbol{P}'}$ 的后 n_4 项，即 $\boldsymbol{S}''_{w/\boldsymbol{P}'} = [status_{n-n_4+1}, status_{n-n_4+2}, \cdots, status_{now}]$，然后将 $n - n_4$ 置为 n_3，即 $\boldsymbol{S}''_{w/\boldsymbol{P}'} = [status_{n_3+1}, status_{n_3+2}, \cdots, status_{now}]$。

(5)判断 $\boldsymbol{S}''_{w/\boldsymbol{P}'}$ 是否与 $status\_value_{w_i}$ 相同，若不相同，则继续遍历，即转步骤(3)；否则，转步骤(6)。

(6)只保留 $\boldsymbol{S}''_{w/\boldsymbol{P}'}$ 的最后 1 项，即 $\boldsymbol{S}''_{w/\boldsymbol{P}'} = [status_{now}]$，同时将 n_3 置为 $n-1$，将初始化标志 I_{init} 置为 1，返回电力业务威胁度 $P_t = 0$，匹配成功，过程结束。其中 I_{init} 的初

始值为0,用于标识是否进行一次白名单完全匹配。

(7)若$I_{\text{init}}=0$,则转步骤(8);否则,转步骤(9)。

(8)判断$\boldsymbol{S}''_{\text{w}/P'}$是否为$\text{status_value}_{w_i}$的子集,若是,则返回$P_{\text{t}}=0$,匹配成功,过程结束;否则,继续遍历白名单规则子集中的其他规则,即转步骤(3)。

(9)判断$\boldsymbol{S}''_{\text{w}/P'}$是否与$\text{status_value}_{w_i}$的前$n-n_1$项相同,若相同,则返回$P_{\text{t}}=0$,匹配成功,过程结束;否则,继续遍历白名单规则子集中的其他规则,即转步骤(3)。

1.6 黑白名单相似度匹配

当白名单模式匹配失败后,需要对当前状态序列进行黑白名单相似度匹配,得到相似度P_{g},具体流程如下。

(1)根据$\boldsymbol{P}'$找到对应的控制块黑名单规则子集$U_{\text{b}/P'}=\{\text{status_seq}_{b_1},\text{status_seq}_{b_2},\cdots,\text{status_seq}_{b_t}\}$和白名单规则子集$U_{\text{w}/P'}=\{\text{status_seq}_{w_1},\text{status_seq}_{w_2},\cdots,\text{status_seq}_{w_e}\}$。

(2)对$\forall\text{status_seq}_{b_i}\in U_{\text{b}/P'}$,若$c_{S'}<C_{\text{status_seq}_{b_i}(\text{status_value})}$($c_{S'}$为$\boldsymbol{S}'$的列数,$\text{status_seq}_{b_i}(\text{status_value})$为状态序列$\text{status_seq}_{b_i}$中的status_value值),则有:

$$\begin{aligned}&\text{status\_value}_{\text{status\_seq}'_{b_{i_1}}}=\\&\quad\text{status\_seq}_{b_i}(\text{status\_value})[1,2,\cdots,c_{S'}]\\&\text{status\_value}_{\text{status\_seq}'_{b_{i_2}}}=\\&\quad\text{status\_seq}_{b_i}(\text{status\_value})[2,3,\cdots,c_{S'}+1]\\&\qquad\vdots\\&\text{status\_seq}_{\text{status\_seq}'_{b_{i_k}}}=\\&\quad\text{status\_seq}_{b_i}(\text{status\_value})[c_{\text{status\_seq}_{b_i}}-c_{S'}+1,\\&\quad c_{\text{status\_seq}_{b_i}}-c_{S'}+2,\cdots,c_{\text{status\_seq}_{b_i}}]\end{aligned}$$

否则,$\text{status\_value}_{\text{status\_seq}'_{b_i}}=\text{status\_seq}_{b_i}(\text{status\_value})$。其中,$\text{status_seq}_{b_i}(\text{status_value})[1,2,\cdots,c_{S'}]$为矩阵$\text{status_seq}_{b_i}(\text{status_value})$的第$1—c_{S'}$列。

(3)计算当前状态序列$\text{status_seq}_{\text{now}}$与黑名单中各个规则的距离,计算公式为:

$$d_{\text{b}}(\boldsymbol{v}_{\text{status\_seq}'_{b_i}})=\sum_{i=1}^{c_{\boldsymbol{v}_{\text{status\_seq}'_{b_i}}}}d((\boldsymbol{v}_{\text{status\_seq}'_{b_i}}[\cdot i])^{\text{T}},(\boldsymbol{S}'[\cdot c_{S'}-c_{\boldsymbol{v}_{\text{status\_seq}'_{b_i}}}+i])^{\text{T}})\tag{1}$$

其中,将$\text{status_value}_{\text{status_seq}'_{b_i}}$简记为$\boldsymbol{v}_{\text{status_seq}'_{b_i}}$;$\boldsymbol{v}_{\text{status_seq}'_{b_i}}[\cdot i]$为$\boldsymbol{v}_{\text{status_seq}'_{b_i}}$的第$i$列;函数$d(\boldsymbol{A},\boldsymbol{B})$表示求列向量$\boldsymbol{A}=[x_1,x_2,\cdots,x_n]^{\text{T}}$与$\boldsymbol{B}=[y_1,y_2,\cdots,y_n]^{\text{T}}$的欧几里得距离,即$d(\boldsymbol{A},\boldsymbol{B})=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2}$。

然后计算所有距离的最小值$d_{\text{b_min}}$:

$$d_{\text{b\_min}}=\min\{d_{\text{b}}(\boldsymbol{v}_{\text{status\_seq}'_{b_1}}),d_{\text{b}}(\boldsymbol{v}_{\text{status\_seq}'_{b_2}}),\cdots,d_{\text{b}}(\boldsymbol{v}_{\text{status\_seq}'_{b_t}})\}\tag{2}$$

其中,min{·}表示取最小值。

(4)对$\forall\text{status_seq}_{w_i}\in U_{\text{w}/P'}$,若$c_{S'}<C_{\text{status_seq}_{w_i}(\text{status_value})}$,则有:

$$\begin{aligned}&\text{status\_value}_{\text{status\_seq}'_{w_{i_1}}}=\\&\quad\text{status\_seq}_{w_i}(\text{status\_value})[1,2,\cdots,c_{S'}]\\&\text{status\_value}_{\text{status\_seq}'_{w_{i_2}}}=\\&\quad\text{status\_seq}_{w_i}(\text{status\_value})[2,3,\cdots,c_{S'}+1]\\&\qquad\vdots\\&\text{status\_seq}_{\text{status\_seq}'_{w_{i_k}}}=\\&\quad\text{status\_seq}_{w_i}(\text{status\_value})[c_{\text{status\_seq}_{w_i}}-c_{S'}+1,\\&\quad c_{\text{status\_seq}_{w_i}}-c_{S'}+2,\cdots,c_{\text{status\_seq}_{w_i}}]\end{aligned}$$

否则,$\text{status\_value}_{\text{status\_seq}'_{w_i}}=\text{status\_seq}_{w_i}(\text{status\_value})$。

(5)计算当前状态序列$\text{status_seq}_{\text{now}}$与白名单中各个规则的距离,计算公式为:

$$d_{\text{w}}(\boldsymbol{v}_{\text{status\_seq}'_{w_i}})=\sum_{i=1}^{c_{\boldsymbol{v}_{\text{status\_seq}'_{w_i}}}}d((\boldsymbol{v}_{\text{status\_seq}'_{w_i}}[\cdot i])^{\text{T}},(\boldsymbol{S}'[\cdot c_{S'}-c_{\boldsymbol{v}_{\text{status\_seq}'_{w_i}}}+i])^{\text{T}})\tag{3}$$

其中,将$\text{status_value}_{\text{status_seq}'_{w_i}}$简记为$\boldsymbol{v}_{\text{status_seq}'_{w_i}}$。

然后计算所有距离的最小值$d_{\text{w_min}}$:

$$d_{\text{w\_min}}=\min\{d_{\text{w}}(\boldsymbol{v}_{\text{status\_seq}'_{w_1}}),d_{\text{w}}(\boldsymbol{v}_{\text{status\_seq}'_{w_2}}),\cdots,d_{\text{w}}(\boldsymbol{v}_{\text{status\_seq}'_{w_e}})\}\tag{4}$$

(6)计算相似度P_{g},计算公式为:

$$P_{\text{g}}=d_{\text{w\_min}}/(d_{\text{b\_min}}+d_{\text{w\_min}})\tag{5}$$

1.7 考虑电网时间风险与业务重要性的威胁度修正

(1)电网时间风险。

研究表明,电网在不同时间遭遇恶意入侵的概率存在差异,并且具有统计模式,其统计模式与攻击者的习惯、电力系统的脆弱性等因素相关[7]。本文借用Premaratne U K[7]等人的工作来计算电网时间风险。

Premaratne U K等人根据蜜罐项目收集的实际系统的网络攻击统计数据,即不同时期系统绘制的入侵报告的数量$n(t)$,定义了系统的威胁指数$T(t)$的计算公式[7],如式(6)所示。

$$T(t)=n(t)/\max\{n(t)\}\tag{6}$$

由于人类活动的变化,电力需求在一天内是变化的,且表现出规律性。在一个典型日内,中午的电力需求很高,夜间有一个明显的负荷高峰,之后会下降,直到04:00时刻左右。电力中断将导致所有消费者的损失。因此,可以认为电力系统的脆弱性与输送的电量成比例。为了对消费者造成最大的伤害,攻击者最有可能在高功率需求时攻击系统。据此,Premaratne U K等人定义了系统的脆弱性指数$V(t)$

的计算公式[7]，如式(7)所示。

$$V(t)=P_{\text{avg}}(t)/\max\{P_{\text{avg}}(t)\} \tag{7}$$

其中，$P_{\text{avg}}(t)$为多日t时刻输送的电量平均值。

电网时间风险指数$R(t)$的计算公式[7]为：

$$R(t)=T(t)V(t) \tag{8}$$

根据Premaratne U K[7]等人计算所得电网时间风险指数图可知，一天中时间风险指数最高的时刻为16:00。

(2)业务重要性。

电网中不同业务的重要性各不相同，为了对电网造成最大的伤害，攻击者最有可能对重要性较高的业务进行逻辑篡改。本文中一个业务对应一个控制块，而一个控制块包含多个不同的节点，因此，不同业务的重要性与其关联的各个节点的重要性相关。据此，本文定义业务重要性P_{c}的计算公式为：

$$P_{\text{c}}(\text{pos\_seq})=\frac{\frac{1}{n}\sum_{i=1}^{n}P_{\text{Node}}(\text{pos}_i)}{\max\{P_{\text{Node}}(\text{pos}_t)\}} \tag{9}$$

其中，$P_{\text{Node}}(\text{pos}_i)$为节点$\text{pos}_i$的重要性。

(3)业务威胁度修正。

结合电网时间风险与控制块重要性，对相似度P_{g}进行修正，得到不在黑白名单中当前状态序列的电力业务威胁度P_{t}，计算公式为：

$$P_{\text{t}}=P_{\text{g}}R(t)P_{\text{c}} \tag{10}$$

1.8　威胁度判定

综上，电力业务威胁度P_{t}的计算公式可表示为：

$$P_{\text{t}}=\begin{cases}1 & S'\in U_{\text{b}}\\ P_{\text{g}}R(t)P_{\text{c}} & S'\notin U_{\text{b}}\ \&\ S'\notin U_{\text{w}}\\ 0 & S'\in U_{\text{w}}\end{cases} \tag{11}$$

基于业务威胁度P_{t}，可判断电网测控终端是否遭受电力业务报文攻击。为降低本文所提方法的误报率，定义安全风险阈值X_{safe}，且X_{safe}默认置为0.25。当$P_{\text{t}}>X_{\text{safe}}$时，认为当前状态序列$\text{status_seq}_{\text{now}}=\{S',P'\}$中的$P'$值对应的业务遭受电力业务报文攻击；否则，认为当前电力工控系统处于安全状态，未遭受电力业务报文攻击。

2　电力业务报文攻击识别系统架构

应用上述基于业务逻辑的电力业务报文攻击识别方法的一种系统功能模块如图3所示，其包括捕获模块、预处理模块、深度包解析模块、攻击识别模块与告警模块这5个功能模块。

(1)捕获模块：对电力工控系统传输的报文进行捕获。系统通过实时捕获网络报文流，提取802.3帧数据，保存报文数据，提供报文数据读取接口。

(2)预处理模块：对捕获模块收集的报文进行过滤，仅保留电力业务报文。通过读取报文包头信息判断报文是否为业务报文，避免对非业务报文进行深度包解析，从而提高检测系统的运行效率，如通过检测报文以太网层中ethtype字段是否为0x88B8判断该报文是否为IEC 61850协议中的GOOSE报文，通过检测报文以太网层中ethtype字段是否为0x88BA判断该报文是否为IEC 61850协议中的SV报文，通过检测报文源端口或者目的端口是否为102判断该报文是否为IEC 61850协议中的MMS报文等。

(3)深度包解析模块：对电力业务报文中的业务数据进行提取。为了对电网业务的逻辑状态进行审查，需要获取业务数据，因此需要对业务报文的应用层字段进行解析。本系统采用深度包解析技术提取业务报文应用层数据。

(4)攻击识别模块：基于解析的业务报文数据，利用本文所提网络攻击识别方法，识别电网中是否存在网络攻击。

(5)告警模块：根据攻击识别模块提供的信息，给出相应的告警信息。根据业务地址序列，定位遭受攻击的业务及可能遭受网络攻击的电网测控终端，即负责该业务的电网测控终端，并通过进一步审查，定位最终的遭受攻击的终端。

此外，为了使本文所提网络攻击识别系统能够有效识别电力工控系统可能遭受的网络攻击，需要

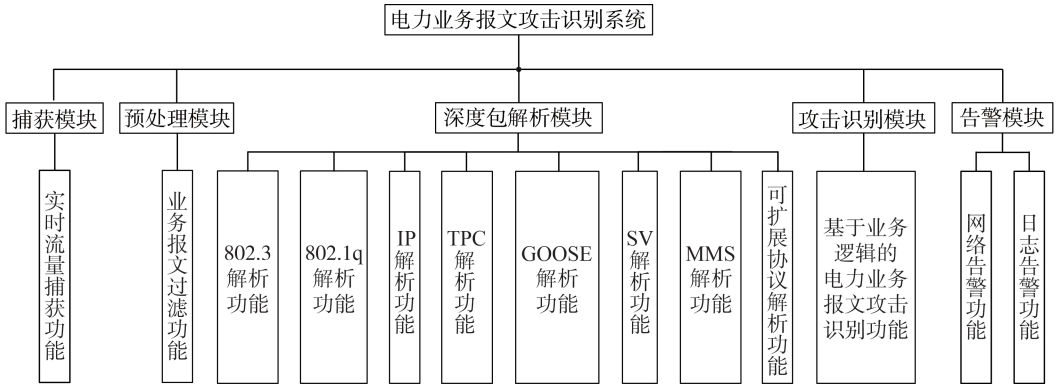

图3　电力业务报文攻击识别系统功能模块

Fig.3　Function modules of PSPA identification system

对黑白名单进行初始化。其中，黑名单中的规则可以通过专家知识进行初始制定，如同一控制块的多个节点同时改变状态；而白名单中的规则不仅可以通过专家知识进行初始制定，还可以通过将黑白名单为空的系统放入电网环境运行48 h后，将系统中收集的所有状态序列作为白名单的初始规则。

3 实验验证与分析

为了验证本文所提网络攻击识别方法能有效地识别电力业务报文攻击，按照图4所示的拓扑结构搭建实验环境，包含GOOSE发布者、GOOSE订阅者、攻击者与网络攻击识别主机。其中GOOSE发布者为智能变电站间隔层智能终端（如测控终端），GOOSE订阅者为智能变电站过程层智能终端（如智能单元），操作一次设备执行发布者发来的控制命令。

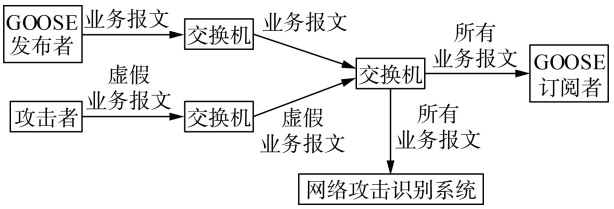

图4 实验环境的拓扑结构

Fig.4 Topology structure of experimental environment

在该环境下，本文通过篡改正常传输的GOOSE报文内容，构造一种电力业务报文攻击。本文用于测试网络攻击识别系统的报文格式模板见附录中图A2，方框中的3个布尔值代表控制块的3个刀闸控制信号，表示该控制块的业务状态。本文通过修改报文中的这3个布尔值来仿真业务逻辑的变化，并通过顺序输入不同状态的业务控制报文来测试识别方法的有效性。

使用packet=(other,appid,0,0,0)对报文进行描述，其中other为其他报文字段，appid为业务号，3个数据对应控制块中3个刀闸信号，即报文alldata中的3个布尔值，取值为0对应实际电力业务中的刀闸开操作，取值为1对应实际电力业务中的刀闸闭操作。

在图4所示实验环境拓扑中设定3个业务，每个业务控制3个刀闸信号，并假设appid为1的业务控制的刀闸信号的重要性为1，appid为2的业务控制的刀闸信号的重要性为2，appid为3的业务控制的刀闸信号的重要性为3。此时，appid为3的业务的重要性最高。

实验中攻击识别系统用的白名单为$U_{w/1}=U_{w/2}=U_{w/3}=\{\text{status_seq}_{w_1}\}$，其中$\text{status_seq}_{w_1}(\text{status_value})=\begin{bmatrix}0&0&0\\0&0&1\\0&1&1\end{bmatrix}$表示该模拟场景下正常的刀闸开合操作序列；用的黑名单为$U_{b/1}=U_{b/2}=U_{b/3}=\{\text{status_seq}_{b_1}\}$，其中$\text{status_seq}_{b_1}(\text{status_value})=\begin{bmatrix}0&1&0\\1&1&0\\1&1&0\end{bmatrix}$表示该模拟场景下一种极其危险的刀闸开合操作序列。需要说明的是，该实验中黑白名单所采用的规则均为假定值，仅用于测试识别方法的有效性，在实际应用中需要结合第2节中所述黑白名单初始化内容，并依据现场环境进行设定。

为了验证黑白名单模式匹配方法的有效性，本文通过篡改正常的报文，构造一种黑名单中违背正常业务逻辑的appid为1的业务报文序列，对报文订阅者发起攻击，该报文序列依次为：

$$\text{packet}_1=(\text{other},1,0,0,0)$$
$$\text{packet}_2=(\text{other},1,0,0,1)$$
$$\text{packet}_3=(\text{other},1,0,1,1)$$
$$\text{packet}_4=(\text{other},1,1,1,1)$$
$$\text{packet}_5=(\text{other},1,0,0,0)$$

在攻击者发送篡改后的报文序列后，该系统发现该报文序列与黑名单中规则status_seq_{b_1}匹配成功，从而识别出网络攻击，并对遭受攻击的业务进行告警，结果见附录中图A3。

为了验证黑白名单相似度匹配方法的有效性，本文模仿攻击者在电网时间风险指数最高的16:00时刻对业务重要度最高的appid为3的业务进行攻击。此时，根据式(9)可知，电网时间风险指数与业务重要性都为1，系统的判定结果由相似度匹配结果决定。通过篡改正常的报文，构造一种未在黑名单中的违背正常业务逻辑的appid为3的业务报文序列，对报文订阅者发起攻击，该报文序列依次为：

$$\text{packet}_1=(\text{other},3,0,1,0)$$
$$\text{packet}_2=(\text{other},3,1,0,1)$$
$$\text{packet}_3=(\text{other},3,0,0,1)$$

在攻击者发送篡改后的报文序列后，该系统发现该报文序列对应的威胁度高于安全阈值，成功识别出网络攻击，并对遭受攻击的业务进行告警，结果见附录中图A4。

为了进一步验证电网时间风险与业务重要性指标对威胁度修正的有效性，本文模拟一种电网突发情况的处理场景，在08:00时刻对appid为2的业务进行一次紧急操作，该操作的报文序列依次为：

$$\text{packet}_1=(\text{other},2,0,0,1)$$
$$\text{packet}_2=(\text{other},2,1,1,1)$$
$$\text{packet}_3=(\text{other},2,0,1,0)$$

当该系统忽略电网时间风险与业务重要性指标进行攻击识别时，该操作序列对应的威胁度为0.631，高于安全阈值0.25，则系统会将该紧急操作误判为攻击而进行告警。当引入电网时间风险与业务重要性指标后，系统能够有效认定该操作序列为正常操作，结果见附录中图A5，此时该操作序列对

应的电网时间风险指数为0.4，业务重要性为0.667，修正后的威胁度为0.168。

上述实验从黑白名单模式匹配、相似度匹配、考虑电网时间风险与业务重要度的威胁度修正3个方面验证了本文所提方法在识别电力业务报文攻击时的有效性。

4　结论

已有攻击识别方法大多关注流量特征、协议规则或报文不同字段的相关性，而未考虑业务逻辑，难以有效识别电力业务报文攻击。本文提出了一种基于业务逻辑的电力业务报文攻击识别方法，该方法将误用检测与异常检测方法相结合，通过业务逻辑黑白名单的模式匹配实现对在黑白名单中的当前业务状态的误用检测；然后，对不在黑白名单中的状态序列进行相似度匹配，并将电网时间风险与业务重要性融入基于黑白名单的相似度匹配异常检测方法之中，对相似度进行修正，以更真实地评估不在黑白名单中的当前状态序列的威胁度，降低了异常检测方法的误报率，从而实现对电力业务报文攻击的高效、准确识别。应用所提方法实现一个攻击识别原型系统，并对实现后的系统进行了攻击识别测试，证明了所提方法的有效性。

为了更好地识别针对电网测控终端的电力业务报文攻击，后续工作将从完善黑白名单规则方面来进一步提高该方法的有效性，从而为提高电力工控系统的网络攻击检测能力提供有效的参考方法。

附录见本刊网络版（http://www.epae.cn）。

参考文献：

[1] XUE Yusheng, YU Xinghuo. Beyond smart grid: cyber-physical-social system in energy future[J]. Proceedings of the IEEE, 2017, 105(12): 2290-2292.

[2] XU G Y, TAO L M, ZHANG D, et al. Dual relations in physical and cyber space[J]. Chinese Science Bulletin, 2006, 51(1): 121-128.

[3] 刘念，余星火，张建华. 网络协同攻击：乌克兰停电事件的推演与启示[J]. 电力系统自动化，2016，40(6)：144-147.
LIU Nian, YU Xinghuo, ZHANG Jianhua. Coordinated cyber-attack: inference and thinking of incident on Ukrainian Power Grid[J]. Automation of Electric Power Systems, 2016, 40(6): 144-147.

[4] 李中伟，佟为明，金显吉. 智能电网信息安全防御体系与信息安全测试系统构建 乌克兰和以色列国家电网遭受网络攻击事件的思考与启示[J]. 电力系统自动化，2016，40(8)：147-151.
LI Zhongwei, TONG Weiming, JIN Xianji. Construction of cyber security defense hierarchy and cyber security testing system of smart grid: thinking and enlightenment for network attack events to national power grid of Ukraine and Israel[J]. Automation of Electric Power Systems, 2016, 40(8): 147-151.

[5] LOPEZ C, SARGOLZAEI A, SANTANA H, et al. Smart grid cyber security: an overview of threats and countermeasures[J]. Journal of Energy and Power Engineering, 2015, 9(7): 632-647.

[6] 张盛杰，顾昊旻，李祉岐，等. 电力工业控制系统信息安全风险分析与应对方案[J]. 电力信息与通信技术，2017，15(4)：96-102.
ZHANG Shengjie, GU Haomin, LI Zhiqi, et al. The information security risk analysis and response plan for power industry control system[J]. Electric Power Information and Communication Technology, 2017, 15(4): 96-102.

[7] PREMARATNE U K, SAMARABANDU J, SIDHU T S, et al. An intrusion detection system for IEC 61850 automated substations[J]. IEEE Transactions on Power Delivery, 2010, 25(4): 2376-2383.

[8] CHEUNG S, DUTERTRE B, FONG M, et al. Using model-based intrusion detection for SCADA networks[J]. Proceedings of the Scada Security Scientific Symposium, 2006.

[9] YANG Y, LITTLER T, WANG H F, et al. Rule-based intrusion detection system for SCADA networks[C]//2nd IET Renewable Power Generation Conference(RPG 2013). Beijing, China: IET, 2013: 1-4.

[10] PARTHASARATHY S, KUNDUR D. Bloom filter based intrusion detection for smart grid SCADA[C]//2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). Montreal, QC, USA: IEEE, 2012: 1-6.

[11] NIVETHAN J, PAPA M. Dynamic rule generation for SCADA intrusion detection[C]//2016 IEEE Symposium on Technologies for Homeland Security(HST). Waltham, MA, USA: IEEE, 2016: 1-5.

[12] KWON Y, KIM H K, LIM Y H, et al. A behavior-based intrusion detection technique for smart grid infrastructure[C]// 2015 IEEE Eindhoven PowerTech. Eindhoven, Netherlands: IEEE, 2015: 1-6.

[13] YANG Y, MCLAUGHLIN K, LITTLER T, et al. Intrusion detection system for IEC 60870-5-104 based SCADA networks [C]//2013 IEEE Power & Energy Society General Meeting. Vancouver, BC, Canada: IEEE, 2013: 1-5.

[14] YANG Y, XU H Q, GAO L, et al. Multidimensional intrusion detection system for IEC 61850-based SCADA networks[J]. IEEE Transactions on Power Delivery, 2017, 32(2): 1068-1078.

[15] HONG J, LIU C C, GOVINDARASU M. Detection of cyber intrusions using network-based multicast messages for substation automation[C]//ISGT 2014. Washington, DC, USA: IEEE, 2014: 1-5.

[16] WU S S, LIU C C, SHOSHA A F, et al. Cyber security and information protection in a smart grid environment[J]. IFAC Proceedings Volumes, 2011, 44(1): 13696-13704.

[17] HONG J, LIU C C, GOVINDARASU M. Integrated anomaly detection for cyber security of the substations[J]. IEEE Transactions on Smart Grid, 2014, 5(4): 1643-1653.

[18] KLEINMANN A, WOOL A. A statechart-based anomaly detection model for multi-threaded SCADA systems[C]//International Conference on Critical Information Infrastructures Security. [S.l.]: Springer International Publishing, 2016: 132-144.

作者简介：

王海翔

王海翔（1994—），男，山东聊城人，硕士研究生，主要研究方向为电力大数据、电力工控安全（**E-mail**：a6215916@qq.com）；

朱朝阳（1974—），男，北京人，教授级高级工程师，博士，主要研究方向为电力信息系统、电力大数据、电力工控安全（**E-mail**：zhucy@epri.sgcc.com.cn）；

王　宇（1996—），男，安徽寿县人，硕士研究生，主要研究方向为电力工控安全（**E-mail**：564943524@qq.com）。

（编辑　陆丹）

Identification method of power service packet attacks based on service logic

WANG Haixiang[1], ZHU Chaoyang[1], WANG Yu[2], ZHANG Ruiwen[2], LI Jun'e[2], LI Jiyuan[3], YING Huan[1]
(1. Information & Communication Department, China Electric Power Research Institute, Beijing 100192, China;
2. Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China;
3. State Grid Zhejiang Electric Power Research Institute, Hangzhou 310014, China)

Abstract: The PSPAs (Power Service Packet Attacks) of power grid measurement and control terminals are easy to cause misoperation of primary electric equipment, thus causing electric power accidents. PSPAs usually achieve the attack purpose by interfering the normal service logic. Existing attack identification methods do not take service logic into account and have poor effectiveness. Therefore, an identification method of PSPAs based on service logic is proposed. The state chain of power service logic, blacklist and whitelist are defined. The misuse detection method and anomaly detection method are combined to evaluate the threat degree of power service based on the service logic blacklist and whitelist. Considering the time risk and service importance of power grid, the threat degree is corrected, and the effective and accurate identification of PSPAs is realized by comparing the service threat degree and the security threshold. The architecture of an attack identification system based on the proposed method is presented, and the system is tested to verify the effectiveness of the proposed method.

Key words: power service packet attacks; identification method of attacks; service logic; state chain; power grid measurement and control terminals

(上接第196页 continued from page 196)

Mechanical fault sound diagnosis based on GFCC and random forest optimized by whale algorithm for dry type transformer

GENG Qishen[1], WANG Fenghua[2], JIN Xiao[3]
(1. College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China;
2. Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;
3. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: To effectively extract the mechanical condition information of power transformer and then identify the mechanical faults via acoustic signals, a mechanical fault sound diagnosis method based on GFCC (Gammatone Filter Cepstral Coefficient) and random forest optimized by whale algorithm for transformer is proposed according to the excellent sound recognition ability of human auditory system. Firstly, the GFCCs of transformer acoustic signal are calculated, and the information entropy is introduced to extract the main acoustic feature information in GFCC. Then the whale algorithm is used to optimize the scale and feature subset of decision tree-based classifier in random forest and the classification model of typical mechanical fault based on optimized random forest is constructed. The calculative results of acoustic signals of a 10 kV dry-type transformer under normal condition and typical mechanical faults show that the typical mechanical fault model of transformer based on GFCC main characteristic parameters and random forest optimized by whale algorithm has better recognition ability with high accuracy of more than 95%, and has excellent anti-noise performance and robustness.

Key words: power transformers; sound signal; fault diagnosis; Gammatone filter cepstral coefficient; information entropy; whale algorithm; random forest

附　录

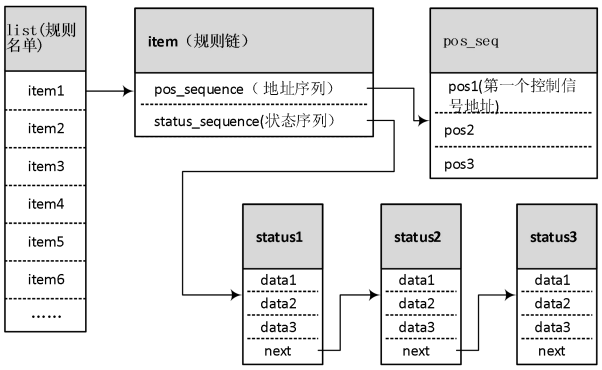

图 A1　黑白名单的数据结构
Fig.A1 Data structure of blacklist and whitelist

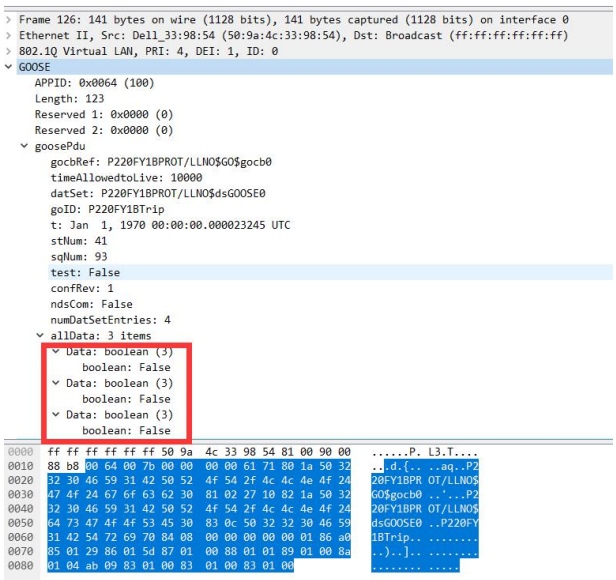

图 A2 GOOSE 报文格式模板
Fig.A2 Format template of GOOSE packet

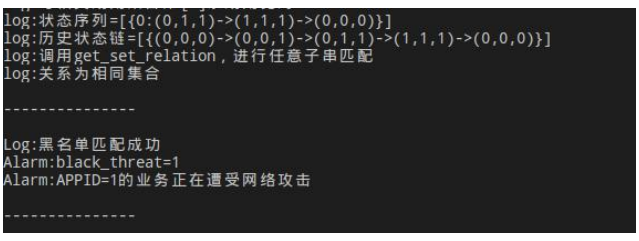

图 A3　黑白名单模式匹配实验结果
Fig.A3 Experimental result of blacklist and whitelist pattern matching

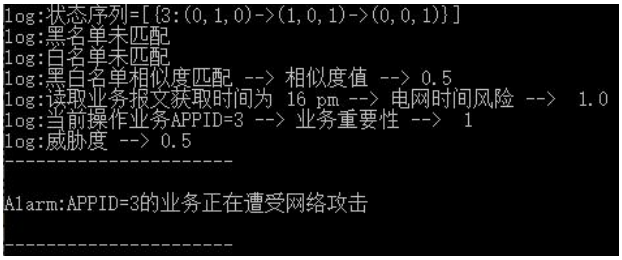

图 A4 相似度匹配实验结果

Fig.A4 Experimental result of similarity matching

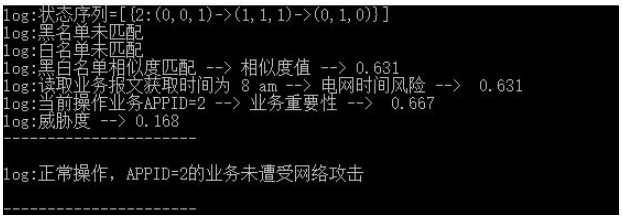

图 A5 考虑电网时间风险与业务重要性的实验结果

Fig.A5 Experimental result considering grid time risk and service importance