「电网」

通用电网工控协议深度解析技术研究

文/东莞供电局　汪　杰　钟志明

首先介绍两种应用最为广泛的通用电网工控协议的特点，即IEC 60870－5协议和IEC 61850协议。然后研究端口识别和深度报文检测两种协议检测方法，进一步分析深度报文检测方法中所用到的特征匹配、字符串多模式匹配和正则表达式匹配3种匹配技术。最后采用深度报文检测技术，对IEC60870－5协议的IEC 104报文和IEC 61850的MMS报文进行举例解析。

随着智能电网的发展，智能电子设备在电网中得到了更为广泛的应用。智能电子设备更多依赖于通信网络进行电网信息的交互，传统电力技术与通信报文结合得更加紧密。电网报文与所采用的通信协议密不可分，电力系统针对自身信息系统的特点，采用了对应的通用电网工控协议。对工业控制协议的通信报文进行深度解析，是电力数据实时去重、协议识别、重组和解析的基础，为电网信息的高效、有效利用提供有力支撑。

早期的电网工控协议主要包括CDT、POLLING、IEC 60870－5协议、IEC 60870－6（TASE.2）协议和各厂家自定义的私有规约。随着信息技术的发展和电网信息需求的增大，电网的信息容量猛增，早期的电网通信方式和协议无法满足电网统一的通信协议要求，因此提出了IEC 61850协议。以当前应用广泛的IEC 60870－5协议和IEC 61850协议为基础，研究通用电网工控协议深度解析技术。

通用电网工控协议

1. IEC 60870－5协议

IEC 60870－5协议是IEC－TC57为适应电力系统传输远动信息的需要而制定的，该协议规定了传输帧格式、链路传输规则、应用数据的一般结构、应用数据的定义和编码以及基本应用功能等主要内容，报文结构如图1所示。

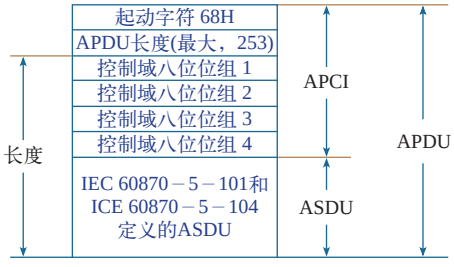

图1　IEC 60870－5报文结构

IEC 60870－5协议具有丰富的内容，是一个结合信息技术逐步发展的通用电网工控协议。1995年，IEC 60870－5－101定义了基本远动任务配套标准；1996年，IEC 60870－5－102定义了电力系统中传输电能脉冲计数量的配套标准；1997年，IEC 60870－5－103定义了继电保护设备信息接口配套标准；2000年，IEC 60870－5－104定义了IEC 60870－5－101协议的以太网传输方式。

其中最主要的协议是IEC 60870－5－104（IEC 104），主要用在远程的四遥（遥信、遥

测、遥控和遥调）等方面，如主网和配网调度自动化系统中的远动控制等。

IEC 60870－5协议具有一系列优点。

1）采用ISO/OSI七层协议简化后的三层增强功能模型（物理层、链路层和应用层），具有更快的响应速度。

2）提供报文的优先级服务。

3）报文灵活性高，适用于多种网络拓扑结构。

但是，在推广IEC 60870－5系列标准的过程中也遇到一些问题。目前工程中所使用的基于TCP/IP的IEC 60870－5－104是面向字符的远动协议，仅仅考虑具体设备的数据格式的统一，并没有对变电站自动化系统内部各种实际对象进行统一的建模和描述，各个厂商按照各自不同的理解进行设计与实现，使得不同系统之间互操作性比较差。

2. IEC 61850协议

针对IEC 60870－5协议的数据模型松散、互操作性差等问题，通用电网工控协议发展了IEC 61850协议。IEC 61850协议采用面向对象的设计思路，满足设备的即插即用，支持从设备到应用进程的自我描述和自我诊断，保证了来自不同制造商的IED之间高度的互操作性。报文结构如图2所示。

IEC 61850协议由早期的变电站自动化领域的协议，发展到当前涵盖整个电力系统自动化的协议，成为了电力系统自动化领域唯一的全球通用标准。IEC 61850提出了一种公共的通信标准，通过对设备的一系列规范化，使其形成一个规范的输出，实现电力信息系统的无缝连接和工程运作标准化。

IEC 61850协议包含的主要报文包括采样值SV报文（当前电力系统的主流SV报文是IEC 61850－9－2格式），以及面向通用对象的变电站事件GOOSE报文、SNTP 对时报文、通用变电站事件GSSE报文和抽象通信服务接口映射到制造报文规范的MMS报文。应用最为广泛的是传递电流和电压量的SV报文、传统开关量的GOOSE报文和电力管理报文MMS，SV报文和GOOSE报文采用直接映射到数据链路层的做法，避开了网络层和传输层，报文结构相对简单。而MMS报文不仅包含了完整的TCP/IP协议，而且涉及了制造报文规范，解析较为复杂，将重点研究MMS报文的解析。

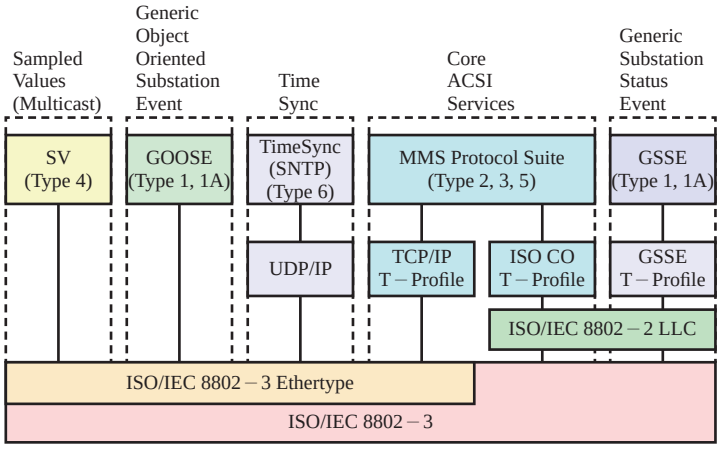

图2　IEC 61850报文

协议解析方法

1. 检测方法

作为早期的流量识别技术，端口识别技术得到了广泛应用。由于绝大部分应用通信是按照端口映射方式实现的，即使用TCP或UDP固定端口进行流量传输，所以根据各个应用协议在IANA（The Internet Assigned Numbers Authority）中注册的端口号来区分不同协议，如MMS协议使用102端口等。端口识别技术利用端口映射特性实现报文检查，实现较为容易。

但是随着应用的快速发展，应用通信开始使用动态端口技术或伪装成常用端口进行传输，利用端口识别工控协议检测准确度降低。

为解决基于端口的识别技术产生的问题，有研究提出了基于载荷的协议识别算法，即从网络数据包中提取用于识别协议的信息。

深度报文检测（Deep Packet Inspection，DPI）技术就是在分析包头的基础上，增加了对应用层负载的分析，是一种基于应用层的流量检测和控制技术，如图3所示。当数据流流经监控设备时，DPI引擎通过匹配特征字、包长等信息对应用层协议进行识别。

图3　深度报文检测

不同的应用通常会采用不同的协议进行数据传输，而各种协议都有其特殊的指纹，通过把数据包和这些指纹信息做比较，从而完成流量协议的精确检测。

DPI技术适用于MMS、S7和MODBUS等标准协议，非加密的P2P协议，存在明显特征的Game协议等类别，其适用范围广泛，可实现对非密协议的精准识别，是目前的主流流量识别的核心技术和主要方法。

其中，基于签名的DPI技术，只检测部分载荷内容，以提高检测效率，是目前较成熟的流量识别技术，绝大部分厂商研发的协议识别与流量检测类产品均使用此类技术。

DPI需要事先详细地分析待识别应用协议，找出不同于其他协议的字段；对于协议特征的提取，若固定字段唯一，则使用该字段作为协议特征串；若存在多个固定字段，则选择其中出现频率最高的特征串来标示协议特征。目前多数此类系统使用正则表达式代替固定的特征串来表示协议特征。

2. 协议匹配技术

（1）特征匹配技术

特征匹配技术作为流量识别，尤其是DPI识别的核心技术，决定了流量识别的匹配效率及准确度。目前特征匹配技术主要包括模式匹配和正则表达式匹配。

模式匹配是数据结构中字符串的一种基本运算；给定一个或多个子串，要求在某个字符串中找出与子串相同的所有子串，即为模式匹配。正则表达式匹配是找出文本T中所有属于给定正则表达式（Regular Expression，RE）的语言L（RE）的子串。下面就模式匹配和正则表达式匹配技术进行阐述。

（2）字符串多模式匹配

所谓多模式匹配，是在文本串$T[1，\cdots，n]$中一次匹配多个模式串P_1，P_2，…，P_k，其中k为模式串个数。$k=1$时，为单模式匹配。多模式匹配比多个模式串逐个进行传统单模式匹配的速度快得多，因此在很多领域都有重要应用，如拼写检查、语言翻译、搜索引擎、网络入侵检测和病毒特征码匹配等。

AC算法是Alfred V.Aho和Margaret J.Corasick于1974年提出（与KMP算法同年）的一个经典的多模式匹配算法，可以保证对于给定的长度为n的文本和模式集合$P\{p_1，p_2，\cdots，p_m\}$，在$O(n)$时间复杂度内，找到文本中的所有目标模式，而与模式集合的规模m无关。

采用AC算法进行匹配需先对模式串集合进行预处理，形成树型有限状态自动机；然后只需对文本串扫描1次，就可找出与其匹配的所有模式串。

在预处理阶段，AC算法会生成goto、failure和output等3个函数，并构建模式集的有限状态自动机；匹配过程从状态机初始状态出发，每次取出目标文本串中的一个字符，使用goto和failure函数进入下一个状态，当某个状态的output函数不为空时，表示在文本串中找到了模式串。

AC算法模式匹配的时间复杂度是$O(n)$，而且与模式集中模式串的个数和每个模式串的长度无关，无论模式串P是否出现在T中，T中的每个字符都必须输入状态机中，所以无论是最好情况还是最坏情况，AC算法模式匹配的时间复杂度都是$O(n)$，包括预处理时间在内，AC算法的总时间复杂度是$O(M+n)$，其中M为所有模式串的长度总和。

（3）正则表达式匹配

正则表达式就是用某种已知模式去匹配一类字符串的一个字符串公式，它主要用来作为描述字符串匹配的工具。正则表达式具有很强的表达能力，它是由一些普通字符（如字符a到z）和一些元字符（metacharacters）组成的，普通字符包括数字和大小写的英文字母，而元字符则有其特殊的含义，正是元字符的存在，才使得正则表达式具有强大的表达能力。不同的解析方式所支持的元字符不尽相同。

正则表达式和自动机密切相关，都是处理字符串问题的重要工具。根据自动机理论和正规文法，两者能够表示完全相同的语言集合，因此可以通过构造有限自动机来识别正则表达式所描述的语言。

有限自动机分为两类，一类是非确定有限自动机（Nondeterministic Finite Automaton，NFA），另一类是确定有限自动机（Deterministic Finite Automaton，DFA）。两者在描述能力上是等价的，它们的主要区别在于，对于确定的状态和输入，NFA允许有多个后继状态，而DFA只有唯一的后继状态。使用自动机来实现正则表达式的匹配时，当自动机的当前状态为一个接受状态时，

就表示识别出了一个模式串。在文本T中搜索正则表达式RE，就是在T中找出所有属于语言L(RE)的子字符串。

利用自动机实现正则表达式匹配的一般过程为首先将正则表达式解析为表达式树，然后转换成NFA，有很多方法可以实现这种转换，但是最常用的方法有两种，一种是Thompson构造法，另一种是Glushkov构造法。

利用NFA直接处理文本时要保存活动状态集合，并且每次读入新的字符时要更新此集合，最坏情况下，匹配过程需要 $O(mn)$的时间，其中m为正则表达式的长度，n为字符的个数，但是其存储空间需求并不大，空间复杂度为$O(m)$。

另一种方法是将NFA确定化，构造相应的DFA，然后利用DFA进行文本匹配。这样在读入字符时可以直接转移，达到$O(n)$的匹配时间。但在最坏情况下，构造 DFA 所需的时间和空间复杂度都是$O(2m)$。

通用电网工控协议深度解析

采用深度报文检测技术，对IEC 104和IEC 61850的MMS这两种应用最广泛的通用电网工控协议进行解析。为了实现IEC 104和MMS报文的深度分析，在分析报文包头的基础上，增加了对应用层负载的分析。采用字符串多模式匹配技术，加强两种报文应用层的流量检测和分析，当两种报文的数据流进入到电网信息系统时，DPI引擎通过匹配特征字、包长等信息对整个IEC 104和IEC 61850报文进行识别。

1. IEC 104报文解析

采用按照IEC 104的协议标准，使用开源代码对IEC 104工控报文的解析，并基于C语言进行解析编码的优化，可达到百兆的线速解析。

依据IEC 104协议规约，对IEC 104报文的字段顺序依次解析，实现IEC 104报文的类型、子类型、指令和变量识别，解析得到源IP、目的IP、源MAC、目的MAC、源端口、目的端口和数据长度等报头信息，并识别得到源设备名（用源IP表示）、目的设备名（用目的IP表示）、传输层协议（TCP）、应用层协议（MODBUS）和报文时间等附加信息。进一步对IEC 104报文的应用层数据进行解析，通过变量阈值的检查和正常行为的建模，得到IEC 104应用层具体内容的深度解析，如图4所示。

源IP	100.100.100.2	目的IP	100.100.100.3:0
源MAC	00:00:54:14:6a:f1	目的MAC	00:1e:ec:0c:d8:86
源端口	502	目的端口	3002
源设备名	100.100.100.2	目的设备名	100.100.100.3:0
传输层协议	TCP	应用层协议	MODBUS
数据长度	13	时间	2013－11－14 10:34:19
数据	{ "transaction id": 132, "protocol id": 0, "len": 7, "unit id": 0, "function_code": 3, "uint8_t_count": 4, "sub_register_value_array": [, { "register_value": 0 }, { "register_value": 0 },] }		

图4 IEC 104协议深度解析

2. IEC 61850的MMS报文解析

对于MMS等采用抽象语法记法（ASN.1）来描述APDU的工控以太网协议，利用基本编码规则实现ASN.1描述的抽象语法和比特流形式的传输语法的转换，因此如何实现ASN.1解码成了该类型协议解析的关键。本方法解决的技术问题所采用的技术方案是基于使用工业协议的ASN1语法模板，使用ASN1代码转换模块自动生成针对该协议的解析代码，并实现各个工控协议解析的统一格式输出，达到对于该种类型工控协议的规范化快速解析，得到MMS协议的最终解析效果，如图5所示。

源IP	10.0.0.200	目的IP	10.0.0.201
源MAC	e4:11:5b:3e:dd:90	目的Mac	3c:97:0e:6f:19:6a
源设备ID	58b453f6847b61165434aa8c	目的设备ID	58b453d4847b61165434aa71
源设备名称	10.0.0.200	目的设备名称	10.0.0.201
应用层协议	MMS	时间	2017－02－28 11:17:02
报文类型	confirmed－ResponsePDU	操作指令	read
数据	{"confirmed-ResponsePDU":{"Confirmed-ResponsePDU":{"invokeID":287,"read":{"read-Response":{"listOfAccessResult":[{"success":{"structure":{"structure":[{"integer":{"INTEGER":1}},{"bit-string":{"BIT_STRING":[0,0]}},{"utc-time":{"UtcTime":[88,-76,-21,-96,-106,-53,-64,0]}}]}}}}}}}}}		

图5 MMS报文深度解析

结束语

随着信息技术更多应用于电网，电网更多以通信网络的形式传递信息，对通用电网工控协议的解析显得愈发重要。首先分析了IEC 60870－5和IEC 61850两种电网工控协议的报文特点，接着研究当前信息领域主要的协议深度解析方法，最后以IEC 60870－5的IEC 104报文和IEC 61850的MMS报文为例，介绍深度报文检测技术对这两种报文的解析情况。**EA**

（收稿日期：2018.04.08）

（编辑：刘群垚）