基于 SIMI 模型的 S7 协议的实时异常流量检测方法

陈 曦[1,2],姜亚光[2],李建彬[3],闫靖晨[3],刘曙元[4],李坤昌[3]
(1.北京大学 软件与微电子学院,北京 102600;2.中国软件评测中心,北京 100044;
3.华北电力大学 控制与计算机工程学院,北京 100026;
4.国家能源集团华电天仁电力控制技术有限公司,北京 100039)

摘 要:S7 协议在通信过程中存在着脆弱性,使得工业生产通信过程容易受到攻击,造成极大的安全隐患。为了解决这个问题,提出一种基于 SIMI 的 S7 协议异常流量检测方法。首先分析了 S7 协议的特征以及脆弱性;然后,提出一种基于 SIMI的 S7 协议实时异常流量检测方法,该方法在流量状态特征关联性分析的基础上,利用 SIMI 算法的分类特性对 S7 协议异常流量实时状态进行有效识别和分类,构建出 S7 协议异常流量状态的知识图谱;最后,通过模拟实验验证了方法的有效性。通过分析,算法的计算复杂度显著降低。

关键词:S7 协议;异常流量检测;脆弱性;SIMI
中图分类号:TN918.91 **文献标识码**:A DOI:10.16157/j.issn.0258-7998.191316

中文引用格式:陈曦,姜亚光,李建彬,等. 基于 SIMI 模型的 S7 协议的实时异常流量检测方法[J].电子技术应用,2020,46(8):101-106.
英文引用格式:Chen Xi,Jiang Yaguang,Li Jianbin,et al. A real time abnormal traffic detection method based on SIMI model for S7 protocol[J]. Application of Electronic Technique,2020,46(8):101-106.

A real time abnormal traffic detection method based on SIMI model for S7 protocol

Chen Xi[1,2],Jiang Yaguang[2],Li Jianbin[3],Yan Jingchen[3],Liu Shuyuan[4],Li Kunchang[3]
(1.School of Software & Microelectronics,Peking University,Beijing 102600,China;
2.Department of CSTC in Network Security Inspection and Evaluation of Industrial Control System,Beijing 100044,China;
3.School of Control and Computer Engineering,North China Electric Power University,Beijing 100026,China;
4.Beijing Huadian TianRen Electric Power Control Technology Company Limited,Beijing 100039,China)

Abstract:The vulnerability of the S7 protocol in the communication process makes the industrial production communication process vulnerable to attacks, causing great security risks. To solve this problem, this paper proposes a method for detecting abnormal traffic of the S7 protocol based on SIMI. Firstly, the characteristics and vulnerability of the S7 protocol are analyzed. Then, a real-time abnormal traffic detection method of the S7 protocol based on SIMI is proposed. Based on the correlation analysis of traffic status characteristics, this method uses the classification characteristics of the SIMI algorithm to effectively identify and classify the real-time status of abnormal traffic in the S7 protocol, and build a knowledge map of the abnormal traffic status of the S7 protocol. Finally, the validity of the method is verified by simulation experiments. Through analysis, the computational complexity of the algorithm is significantly reduced.

Key words:S7 protocol;abnormal traffic detection;vulnerability;SIMI

0 引言

随着工业化与信息化进程的不断交叉融合,越来越多的信息技术应用到了工业控制领域。其中数据交换是依靠管理信息与生产控制网络两者之间的交互实现,这样的方式使得工业控制系统与各种管理系统紧密联系,互相通信,而不再像以往一样是一个独立运行的系统[1]。随着时间的推移,从搜集远程设备信息的简单网络到内置冗余设备的复杂系统网络,工业控制网络通信协议(工控协议)的发展进程从未停歇,而且工控系统内部所使用的协议有时只是针对某个特定的应用程序[2-3]。

大多数工控协议都有一个共同点,那就是这些协议在设计之初都没有考虑到随之而来的安全问题,因此这些协议与生俱来就不安全。自从工控协议与 IT 网络相融合以及主流工控协议开始基于 TCP/IP 协议[4]构建以来,安全性就成为了工业控制系统的一个重要问题[5]。它面临着大量协议数据明文传输,缺乏认证和加密,存在被窃听、伪装、篡改、抵赖和重放攻击等风险[6]。如 2010 年伊朗“震网”病毒事件、2015 年乌克兰电网攻击

事件的巨大影响，进一步表明工控行业的相关安全问题已经上升到了国家安全的高度[7]。类似的事件无一不是造成巨大的影响，攻击者利用一系列的漏洞和手段，入侵了对方的控制系统，而最终都是通过控制器直接操作了相应的PLC，这些操作都承载在对应的工控协议上[8]。

作为专有工控协议，西门子S7协议在电力系统的应用十分广泛[9-10]。其应用主要有化学领域中的水处理、气力除灰、(特)高压直流输电系统中的应用[11]。在这些环节中，采用PLC后热电厂中每个环节都会得到很大的提高[12]。异常流量检测技术在S7协议中应用广泛。基于流量的异常检测技术是根据外部入侵和内部破坏等攻击行为的流量特征，在各个网络连接链路采集数据进而实现对入侵行为的检测[13]。STAVROULAKIS P等人详细讨论了工控系统入侵检测技术的流量分析方法[14]。VOLLMER T等人采用BP神经网络来训练网络流量模型实现入侵检测[15-16]。

本文提出一种基于SIMI的S7协议异常流量检测方法。首先，在模拟仿真环境中通过Wireshark获取S7协议的流量包，分析并提取流量包中流量的特征以及验证协议脆弱性；然后，基于SIMI构建S7协议异常流量状态的知识图谱。该方法在流量的状态特征的关联性分析的基础上，利用SIMI算法的分类特性对S7协议异常流量状态进行有效识别和分类，进而构建S7协议异常流量状态的知识图谱。最后，通过实验验证了本文提出方法的有效性。另外，实验表明算法的计算复杂度从$O(n^2)$降到$O(n)$。该方法采用分片的相似度分析(Similarity，简称为SIMI)方法构建异常模型，与前人的研究方法不同的是，以往的方法是在采集到的整个帧做分类或者聚类，没有考虑到实际流量传输过程中的数据传输包含了很多冗余的数据，增加了整个模型的计算复杂度和时间成本，而本文提出的模型会在预处理阶段对帧中包含的各个字段的含义进行分析，去除了对计算无意义的冗余数据，降低了计算维度并最终降低整个模型的计算复杂度。与神经网络方法相比，SIMI算法更适合实时性要求较高的工业控制系统。

1 S7协议

S7通信协议，或称之为Step 7通信协议，由Siemens公司基于某ISO协议实现，该协议在各个工控行业，尤其是电力系统中都有广泛的应用。S7协议是西门子的内部的不对外公开的协议，其协议结构如图1所示。

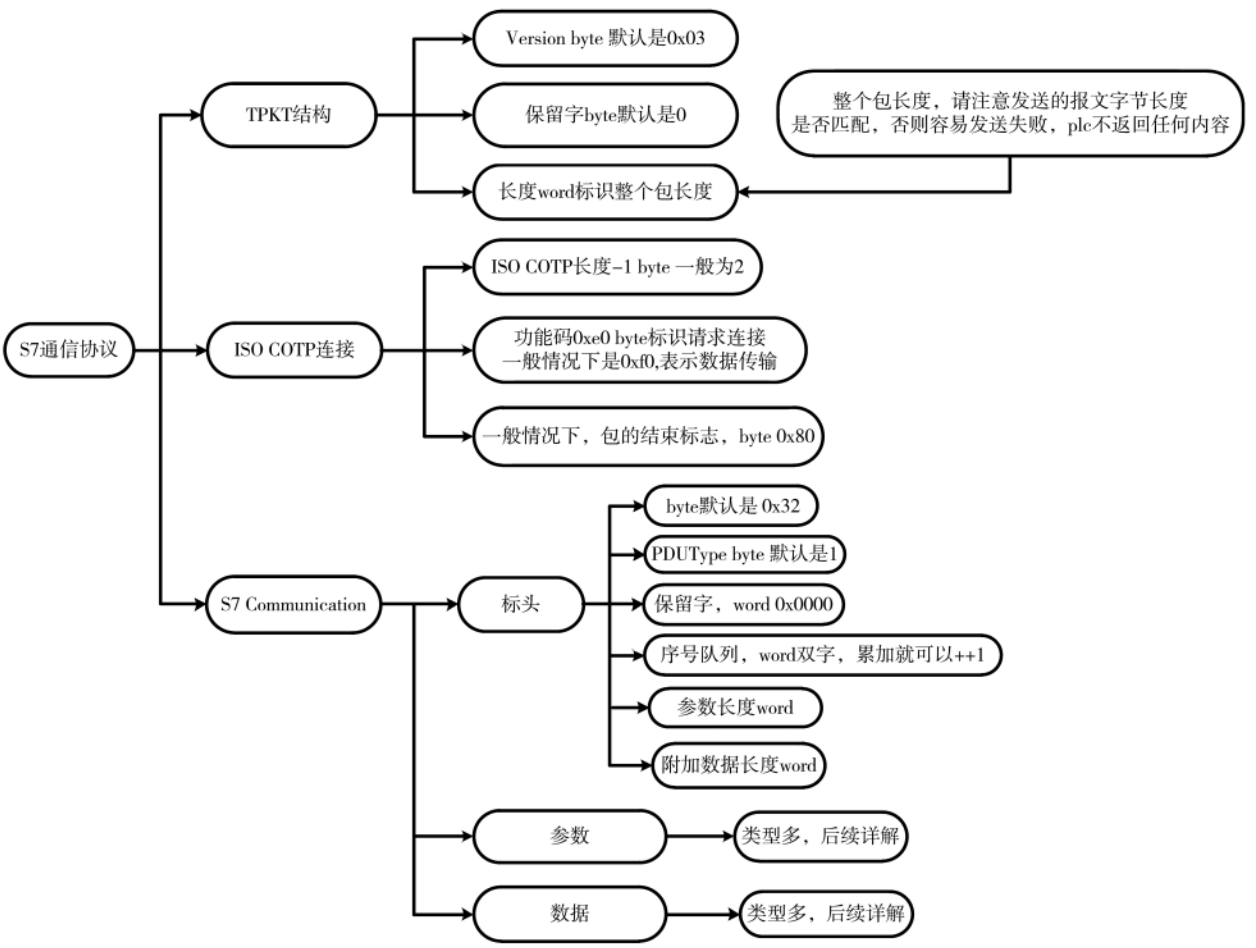

图1 S7协议结构示意图

1.1 S7 协议详解

S7 是西门子的专有通信协议，它通过以太网在 S7 系列的 PLC 之间进行通信，可用于 PLC 编程、数据交换、诊断目的以及 SCADA 系统的 PLC 数据访问。表 1 为 S7 协议的层次模型。

表 1 S7 协议的层次模型

层	OSI	S7 以太网协议模型
7 层	应用层	S7 Communication
6 层	表示层	S7 Communication(COTP)
5 层	会话层	S7 Communication(TPKT)
4 层	传输层	ISO-on-TCP(RFC 1006)
3 层	网络层	IP
2 层	数据链路层	Ethernet
1 层	物理层	Ethernet

S7 协议的 OSI 结构模型如图 2 所示。由图可知，S7 协议中的 TCP/IP 模块主要实现传输服务，并且是基于面向块的 ISO 传输服务。图中的中间部分主要是 TPKT 和 ISO-COTP 协议，其中 S7 协议位于两者之后，该协议允许协议数据单元(PDU)通过 TCP 承载。

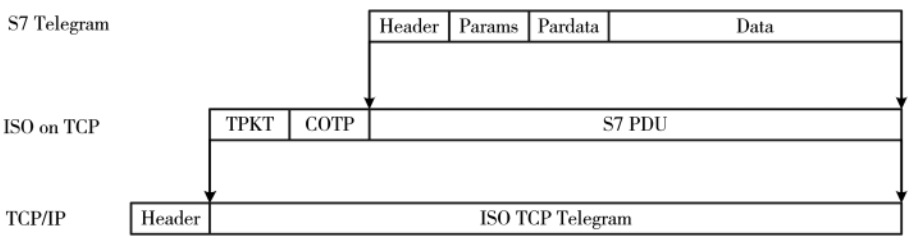

图 2 S7 Comm 协议 OSI 模型

1.2 S7 协议帧格式

S7 PDU 的重要组成部分是标头(Header)、参数(Parameter)、数据(Data)，其中标头主要是关于长度、PDU 参考和消息类型常量的信息；参数部分的内容和结构会根据 PDU 的消息和功能类型而发生变化；数据部分是一个可选字段，例如存储器值、块代码、固件数据等。S7 协议帧格式如图 3 所示。

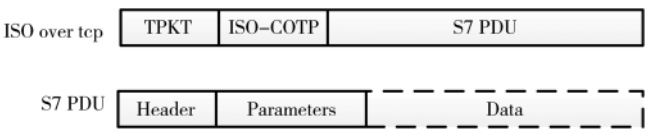

图 3 S7 协议帧格式

S7 Communication 协议的结构随着功能的变化而变化。例如，请求数据报文不包含数据部分。

在建立连接的时候 S7 协议决定并行传输过程中的数量和 PDU 长度的最大为多少。S7 协议传输过程中有请求和应答两部分，这是由于它是面向功能和命令的。

2 基于 SIMI 模型的 S7 协议的实时异常流量检测方法

2.1 相似度计算

对于长为 L 的帧，根据表 1 的 P 个协议参数，从中选取个对当前环境有意义的参数。去重过程：选取有意义的参数需根据当前的环境来定，每个环境中对于参数的偏向是不同的。例如针对 DDoS 攻击的实验环境，需要重点考虑源 IP、目的 IP、源端口、目标端口和协议的时间等参数。此时 S7 协议数据流可以表示为：

$$Q\_P=\{Q_1,Q_2,Q_3,\cdots,Q_p\} \tag{1}$$

其中，Q_P 表示 S7 协议数据流，$Q_1,Q_2,\cdots,Q_p$ 表示数据流中的每一位。去重之后：

$$Q\_P_u=\{Q_{p1},Q_{p2},Q_{p3},\cdots,Q_{pu}\} \tag{2}$$

其中，Q_P_u 表示去重后的 S7 协议数据流，$Q_{p1},Q_{p2},\cdots,Q_{pu}$ 表示去重后数据流中的每一位。以帧长 80 为例：根据上述计算规则：

$$L=80 \tag{3}$$

$$P=26 \tag{4}$$

$$P_u=19 \tag{5}$$

$$Q\_P=\{Q_1,Q_2,Q_3,\cdots,Q_{26}\} \tag{6}$$

$$Q\_P_u=\{Q_{p1},Q_{p2},Q_{p3},\cdots,Q_{19}\} \tag{7}$$

对于每个 S7 协议数据包涉及的帧，每 8 bit 为 1 B，每个字节为一个可以分析的最小数据单元 U。最小数据单元由每两位十六进制的字符组成，为可处理的最小单元。对于每个长度为 L 的帧，包含 L 个最小数据单元。若用 i 表示最小数据单元的顺序，$\forall i\in(0,L)$，上述有意义的参数 P_u 所占的字节数为 K，则每个帧 S 可以表示为：

$$S=\{U_1,U_2,U_3,\cdots,U_i,\cdots,U_k\} \tag{8}$$

在模型中，设计了两个帧的对比，对于每个帧，可以表示为：

$$S_1=\{x_1,x_2,x_3,\cdots,x_i,\cdots,x_k\} \tag{9}$$

$$S_2=\{y_1,y_2,y_3,\cdots,y_i,\cdots,y_k\} \tag{10}$$

以 $L=80$ 为例，$k=62$，则：

$$S_1=\{x_1,x_2,x_3,\cdots,x_i,\cdots,x_{62}\} \tag{11}$$

$$S_2=\{y_1,y_2,y_3,\cdots,y_i,\cdots,y_{62}\} \tag{12}$$

定义 1(相似度的计算)：若 x,y 表示两组帧的最小单元，i 表示最小单元的项，k 表示有意义的参数 P_u 所占的字节数。$\forall x_i,y_i$ 使得每个最小单元的相似度 $\mathrm{SIMI}_i=x_i\oplus y_i$ 成立。

定义 2 (收敛函数)：$\forall x_i,y_i,\forall k$ 使得相似度 $\mathrm{SIMI}=\frac{1}{k}\sum_k \mathrm{SIMI}_i=\frac{1}{k}[(x_1\oplus y_1)+(x_2\oplus y_2)+(x_3\oplus y_3)+\cdots+(x_i\oplus y_i)+\cdots+(x_k\oplus y_k)]=\frac{\sum_k(x_i\oplus y_i)}{k}$ 成立。

2.2 基于 SIMI 模型的 S7 协议实时异常流量检测方法

本文提出的基于 SIMI 模型的 S7 协议实时异常流量检测方法原理和方法流程图如图 4、图 5 所示。

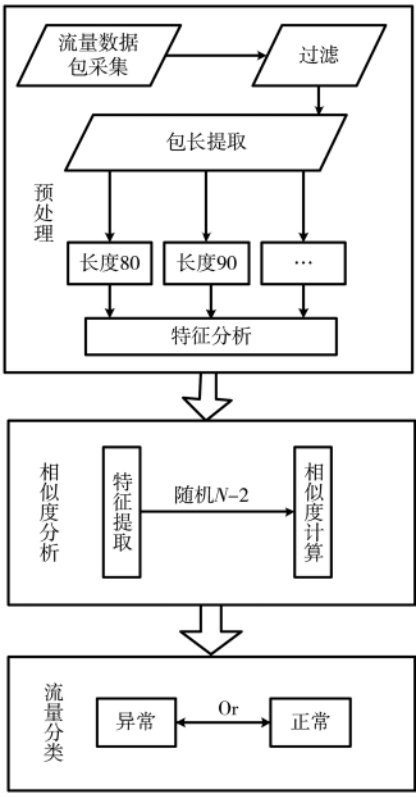

图 4 SIMI 模型模块图

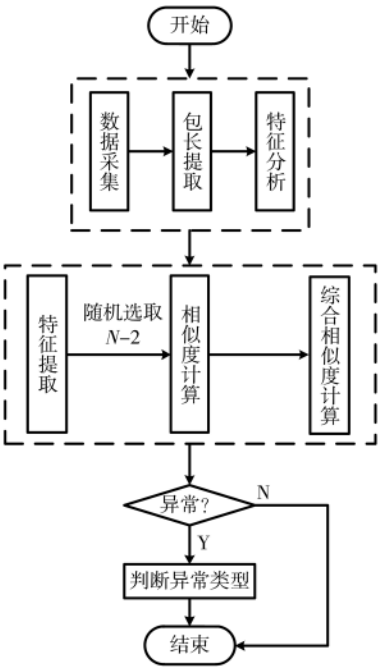

图 5 模型流程图

SIMI 模型的核心部分有预处理、相似度分析、流量分类三部分。预处理阶段：主要对采集到的流量数据包进行过滤、包长提取的过程，不同的包长具有不同的特征，分别对不同的包长进行特征分析；相似度分析阶段：对提取到的相同包长中的 N 个特征样例进行分析，具体过程是随机从特征样例集中选取 $N-2$ 个特征样例进行相似度计算；流量分类阶段：对分析后的结果进行比对，判断流量异常与否。

该模型的预处理阶段主要是对采集到的数据包做数据过滤操作，在以往的异常流量分析[17]过程中所分析的数据是采集到的整个帧做分类或者聚类[18]，没有考虑到实际流量传输过程中的数据传输包含了很多冗余的数据，增加了整个模型的计算复杂度和时间成本[19-20]。而本文模型会在预处理阶段对帧中包含的各个字段的含义进行分析，去除了对计算无意义的冗余数据，降低了计算维度并最终降低整个模型的计算复杂度。

相似度分析模块主要是对预处理过的数据计算其相似度值，该模块是对经过数据过滤之后的每个字段进行分析，以每个字节为处理单元，选取两组相同长度的帧作为相似度计算的依据，利用两组帧的特点采取异或操作，离散计算每个数据单元的相似度值，再计算出每个数据单元的相似度值后用概率论中的平均收敛函数的结果作为整个帧的综合相似度值。对比这两组帧的综合相似度值，选取综合相似度值较高的那组帧作为衡量标准。该相似度值为基准相似度。

流量分析模块主要是利用整个帧的综合相似度值得出流量异常与否的结果。该模块是以相似度分析模块得出的每个帧的综合相似度值为判定依据，综合对比基准相似度与其余 $N-1$ 个帧的综合相似度值，得出 $N-1$ 个对比结果。

假设异常流量为少数情况(小于所有帧数量的一半)，这些对比结果主要分为以下 4 种情况：

(1)若出现 $m(m \leqslant \frac{N-1}{2})$ 个帧的综合相似度值与基准相似度值差距较大，而其余 $N-m-1$ 个帧的综合相似度值与基准相似度值相同或基本相似，则可以认为这 m 个帧属于异常流量。

(2)若出现 $m(m \geqslant \frac{N-1}{2})$ 个帧的综合相似度值与基准相似度值差距较大，而其余 $N-m-1$ 个帧的综合相似度值与基准相似度值相同或基本相似，则可以认为基准帧与剩余的这 $N-m-1$ 个帧属于异常流量。

(3)若出现剩余 $N-1$ 帧的综合相似度值与基准相似度值差距较大，则认为基准帧为异常流量。

(4)若出现剩余 $N-1$ 帧的综合相似度值与基准相似度值相同或基本相似，则认为所有流量均为正常流量。

上述 4 种情况包含了所有流量异常的情形，可以全面判定流量异常与否，其中情况(3)、(4)为情况(1)、(2)的特例。

3 实验验证

3.1 流量数据包采集

本文通过 Wireshark 来分析协议内容，如图 6 所示。

严格执行国家标准 GB3102-93，正确使用量的名称、量的符号与量单位的符号。

对数据包的采集主要在本地搭建的西门子工业控制系统仿真平台中进行，通过人机交互界面组态软件(HMI)实时检测运行情况，在控制端和设备端之间收集数据流并输出为.pcap 文件。将数据包分为正常数据和异常数据，异常数据由对系统进行模拟攻击时产生。本实验 S7 数据包的流量分析情况如图 7 所示。

3.2 数据预处理

数据采集之后过滤无关干扰及非目标协议的数据包，在仿真平台交互数据包中涉及多种帧长，如 80，85，87，93，133，171 等，需要分别对每种帧长进行特征

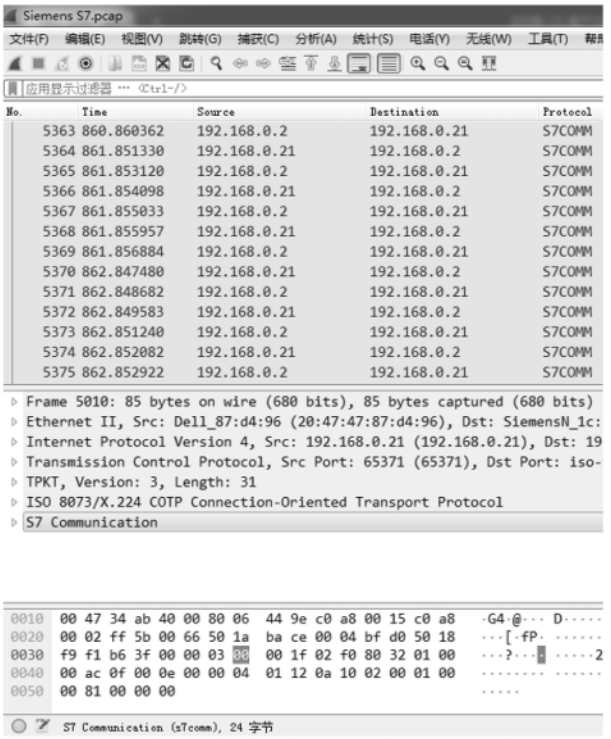

图 6 Wireshark 分析界面

分析及提取。在本文中，将以长度为 80 的帧数据为例，通过对比不同流量数据包中长度均为 80 的帧的相似度，来进行异常流量的检测。由于模拟的上位机和 PLC 是固定的，因此数据的源 IP 和目的 IP、源端口和目的端口应该是一致的。基于此类情况，将一部分与异常相关的字节抽取出来作为新的数据集，最终抽取出来的数据格式和各字段定位如表 2 所示，每单位长度为一个子字段。

3.3 实验结果分析

在数据处理完成后，随机地在所有的 N 条帧数据中抽取两条数据作为检测数据，分别与另外的 $N-1$ 条数据做相似度计算，最终将相似度数值绘图展示，实验结果如图 8 所示。

表 2 数据格式和各字段定位

字段		位置	长度/bit
以太网	目的地址	1-6	6
	源地址	7-12	6
网络协议	差异化服务	16	1
	生命周期	23	1
	协议	24	1
	头部校验	25-26	2
	源 IP	27-30	4
	目的 IP	31-34	4
传输控制协议	源端口	35-36	2
传输控制协议	目的端口	37-38	2
	确认号	43-46	4
	标志位	47-48	2
	校验和	51-52	2
	紧急指针	53-54	2
TPKT		55-58	4
ISO 8073/X.224		59-61	3
S7 协议	头部	62-73	12
	参数	74-75	2

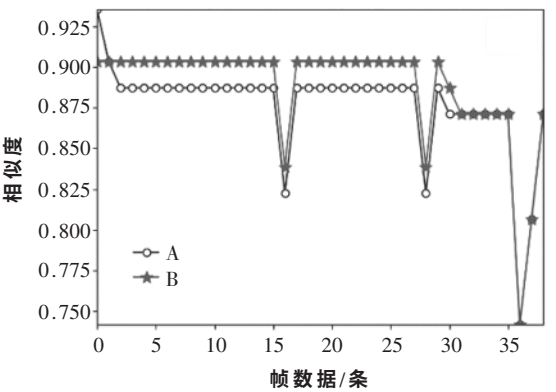

图 8 相似度数值

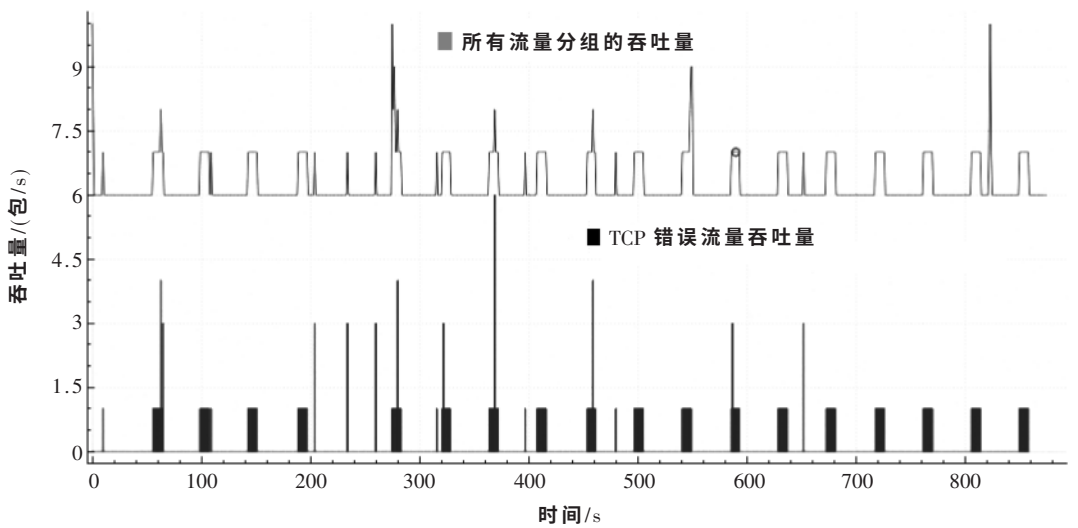

图 7 数据包流量分析

由图 8 可知，二者与其他帧数据的相似度整体相差不大，且对于异常数据的检测结果一致，因此判断这两条检测数据均为正常流量数据，并检测出了 4 条异常流量数据，且定位在了第 18、30、39、40 4 条数据上。对比另一个数据结果如图 9 所示。

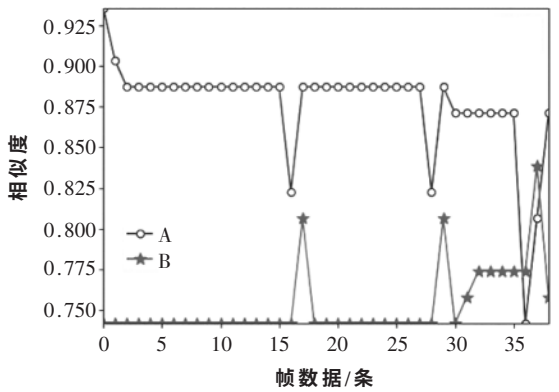

图 9 实验结果对比

由图 9 可知，二者相似度曲线相差较大，且 B 组相似度整体偏低。由此可判断，A 组检测数据为正常流量数据，B 组检测数据为异常流量数据，故通过 A 组检测数据检测出 4 条异常流量数据，而 B 组数据由于本身为异常数据，故不具有检测异常的特性。

对计算复杂度进行分析。通过分析得出，以前的计算复杂度均重复计算了所有文本信息，其计算杂度为 $O(n^2)$，而本文提出的基于 SIMI 模型的 S7 协议实时异常流量检测方法大大去除了冗余信息的计算，降低了计算复杂度，其计算复杂度为 $O(2(n-2))=O(n)$。

4 结论

本文提出了一种基于 SIMI 的 S7 协议异常流量检测方法。首先，分析 S7 协议流量包中流量的特征。然后，提出了一种基于 SIMI 的 S7 协议异常流量检测方法。该方法在流量的状态特征的关联性分析的基础上，利用 SIMI算法的分类特性对 S7 协议异常流量状态进行有效识别和分类。最后，模拟了 S7 协议的实验。通过分析，证明了该算法的有效性，且算法的计算复杂度从 $O(n^2)$降到 $O(n)$。

参考文献

[1] 应欢，刘松华，韩丽芳，等.电力工业控制系统安全技术综述[J].电力信息与通信技术，2018，16(3)：56-63.

[2] 程必成，刘仁辉，赵云飞，等.非标工业控制协议格式逆向方法研究[J].电子技术应用，2018，44(4)：126-129.

[3] 魏骁，刘仁辉，许凤凯.基于静态二进制分析的工控协议逆向解析[J].电子技术应用，2018，44(3)：126-130.

[4] 李瑾博.基于 TCP/IP 协议的工业控制网络远程数据通信网关[J].电子技术与软件工程，2017(17)：13.

[5] 丰大军，张晓莉，杜文玉，等.安全可信工业控制系统构建方案[J].电子技术应用，2017，43(10)：74-77.

[6] KNOWLES W，PRINCE D，HUTCHISON D，et al.A survey of cyber security management in industrial control systems[J]. International Journal of Critical Infrastructure Protection，2015，9：52-80.

[7] 李林峰，房志奇，康卫，等.工业控制防危系统专家规则的管理[J].电子技术应用，2015，41(7)：111-113.

[8] 谢丰.工业控制系统网络攻击与安全防护思考[J].保密科学技术，2016(9)：18-21.

[9] KLEINMANN A，WOOL A.Accurate modeling of the Siemens S7 SCADA protocol for intrusion detection and digital forensics[C].Journal of Digital Forensics，Security and Law，2014.

[10] JU M L，ZHAI X Q，ZHANG Q.Application of siemens S7-300 PLC in the thermal power plant flue gas desulfurization control system[J].Applied Mechanics and Materials，2014，511-512：1123-1127.

[11] BAYINDIR R，CETINCEVIZ Y.A water pumping control system with a programmable logic controller(PLC) and industrial wireless modules for industrial plants-An experimental setup[J].ISA Transactions，2011，50(2)：321-328.

[12] 柴永强.西门子 S7-400H PLC 控制系统在气力除灰系统中的应用[J].科技与企业，2012(16)：129.

[13] 吴军，张向丽，张轶君，等.一种基于 xenVMI 机制下的蜜网流量异常检测方法[J].电子技术应用，2015，41(1)：122-128.

[14] STAVROULAKIS P.Handbook of information and communication security[M].Berlin Heidelberg：Springer，2010.

[15] VOLLMER T，MANIC M.Computationally efficient neural network intrusion security awareness[C].International Symposium on Resilient Control Systems.IEEE，2009.

[16] LINDA O，VOLLMER T，MANIC M.Neural network based intrusion detection system for critical infrastructures[C].Atlanta，Georgia，USA：International Joint Conference on Neural Networks，IJCNN 2009，2009.

[17] 徐玉华，孙知信.软件定义网络中的异常流量检测研究进展[J/OL].软件学报，2019：1-25[2019-11-25].https://doi.org/10.13328/j.cnki.jos.005879.

[18] 刘慕娴，陈文迪，刘桂华.一种基于 K-means 算法的网络流量异常检测模型研究[J].无线互联科技，2019，16(18)：25-27.

[19] 董书琴，张斌.一种面向流量异常检测的概率流抽样方法[J].电子与信息学报，2019，41(6)：1450-1457.

[20] 曾霄笑.基于改进 RNN 及密度聚类的异常流量检测方法[D].北京：北京邮电大学，2019.

(收稿日期：2019-12-03)

作者简介：

陈曦(1987-)，男，硕士，工程师，主要研究方向：工业控制系统质量和安全测评。

姜亚光(1985-)，女，硕士，工程师，主要研究方向：工业控制系统质量和安全测评。

李建彬(1968-)，通信作者，男，硕士，教授，主要研究方向：大数据和信息安全，E-mail：lijb87@ncepu.edu.cn。

版权声明