Hans Journal of Data Mining 数据挖掘, 2017, 7(3), 65-69
Published Online July 2017 in Hans. http://www.hanspub.org/journal/hjdm
https://doi.org/10.12677/hjdm.2017.73007

Hans汉斯

Research on Intrusion Detection Technology Based on Association Rules Mining in Vehicular Networks

Chao Wang, Fei Li

Chengdu University of Information Technology, Chengdu Sichuan
Email: 83448256@qq.com

Received: Jun. 16th, 2017; accepted: Jul. 9th, 2017; published: Jul. 12th, 2017

Abstract

With the development of automobile information, many cars are connected with the external network through the network module. As the car is connected to the extranet, hackers are offered a long way to attack the car via the internet. This paper introduces the background of rough set and association rules, and then uses rough set technology to improve the traditional Apriori algorithm to be applied to the vehicle network intrusion detection, and finally through the test proves the vehicle network intrusion detection.

Keywords

Intrusion Detection, Vehicular Network, Apriori, Association Rules, Rough Set

基于关联规则挖掘的车载网络入侵检测技术研究

王　超，李　飞

成都信息工程大学，四川 成都
Email: 83448256@qq.com

收稿日期：2017年6月16日；录用日期：2017年7月9日；发布日期：2017年7月12日

摘　要

随着汽车信息化，目前很多汽车都通过网络模块与外网连接。由于汽车跟外网相连，那么就给黑客提供

文章引用: 王超, 李飞. 基于关联规则挖掘的车载网络入侵检测技术研究[J]. 数据挖掘, 2017, 7(3): 65-69.
https://doi.org/10.12677/hjdm.2017.73007

了通过网络远程攻击汽车的途径。本文介绍了粗糙集和关联规则的相关背景，接着用粗糙集技术改进传统的Apriori算法应用到车载网络入侵检测方面，最后通过试验验证了对车载网络的入侵检测。

关键词

入侵检测，车载网络，Apriori，关联规则，粗糙集

1. 引言

随着汽车信息化的快速发展，特别是汽车里装置了与外网相连的模块，越来越多的汽车遭遇了入侵威胁，目前已经有了这样的案例[1] [2] [3]。因此，汽车车载网络的安全已经是一个迫在眉睫的问题。目前针对车载网络的入侵检测技术现状是：Groza 通过实验证明在 CAN 总线网络中根据信号特征识别信息发送节点的可行性并推荐其可作为一种 CAN 总线网络的入侵检测方法[4]，Larson 提出了一种建立在安全规则基础上的 CAN 总线网络攻击检测方法，该方法基于 CANopen 协议的对象字典，使用协议级的安全规则检测非法 ECU 行为，并提供了一组示例的安全规则[5]，Muter 给出一种结构化的车载网络异常检测方法，引入一组异常检测传感器对消息帧 ID、数据负载、消息频率等进行检查，并整合传感器结果以防止误报[6]。这些入侵检测方法针对车载网络非关联数据攻击方面检测确实有一定作用，但是没有解决针对车载关联数据攻击方面的入侵检测的问题，所以我把数据挖掘中的关联算法 Apriori 引用到车载网络上，但由于传统的 Apriori 算法效率地下，通过结合粗糙集理论对它进行了改进，主要目的是解决针对车载网络关联数据攻击的问题。

2. 相关背景

2.1. 粗糙集理论

知识在粗糙集理论中被定义为对世界客观事物进行分类的一种能力。在粗糙集理论中，知识必须跟某些所研究对象的整体相联系，这个整体也叫全域或论域。

定义 **1** T = {U,P,V,f}是一个信息系统,其中 U 是被研究对象的集合，P 是研究对象个体的属性集合，V 是属性值域集合。f 是一种映射关系，反应对象集合之间的值。

1) 离散化问题

粗糙集理论相对于德国数学家康托尔提出的集合论来说算是对它的延伸，而处理数据的时候，通常面临数据属性要么是连续性的或者是离散的，但往往处理数据的时候，数据的属性大多数是连续的。因此，属性的连续性转化为离散化问题是它的主要研究内容之一。离散化的优点是可以降低超大数据量。

2) 不完整数据问题

从在数据挖掘的过程中，如果从不完备性的数据集中去挖掘规则会比完备的数据集挖掘规则困难。通常，不完备的数据比较少的情况下，可以删除这些不完备信息，形成一个完整的信息表。但是，如果不完备信息对数据挖掘来说不可缺失的话，就会干扰挖掘结果。而粗糙集理论可以处理这种不完整的信息，并且不需要先验知识。

2.2. 关联规则

设 $I=\{i_1,i_2,\cdots,i_m\}$ 是所研究对象集合中所有属性(项)的集合。其中，如果一个集合里有 k 个属性的属性集就称为 k-项集。设 D 是二维结构的数据集，其中每个记录 T 是项的集合，使得 $T\subset I$ 。设 A 是一个项集，要使记录 T 包含 A，那么只有当且仅当 $A\subset T$ 。

定义 **2** 关联规则的表现形式为 $A\Rightarrow B$ 的蕴含式并且 $A\subset I$， $B\subset I$， $A\cap B=\varnothing$ 。其中 A 称为关联规则的条件，B 称为关联规则的结果。对关联规则 $A\Rightarrow B$ 的度量标准很多：主要有支持度(sup)，置信度(conf)等。定义如下：

$$\text{sup}=(A\Rightarrow B)=P(A\cup B),\ \text{conf}=(A\Rightarrow B)=P(B|A)$$

其中 $P(B)$ 是指 B 在数据集 D 中出现的概率，其它类似。 $\text{sup}(A\Rightarrow B)$ 指 A，B 在 D 中同时出现的概，$\text{conf}(A\Rightarrow B)=P(B|A)$ 表示在 A 存在的情况下，B 存在的条件概率。

3. 粗糙集理论与关联规则挖掘结合

由于原始的 Apriori 算法存在一些不足，因此利用粗糙集理论对 Apriori 做了改进，主要在如下两个方面：1) 利用粗糙集技术对原始数据进行预处理，包括对连续性数据的离散化处理和对不完整数据的补齐。2) 关于在挖掘关联规则过程中通常会形成很多没有意义的规则的问题，采取了对关联规则中属性的制约，使得得出的规则是分类规则，这样就可以处理未知数据。

3.1. 数据的预处理

起初使用粗糙集技术对记录中的条件属性进行预处理，也即对不完备的数据进行补全和对连续性的属性进行离散化处理。这里为了使数据离散化过后仍然保持原始数据的等价关系，选择了 Semi Naive Scaler 离散化算法。

3.2. 属性限制

在关联规则挖掘的实践过程中，有时会遇到得出的规则中包含某些非期望出现的属性。

由于数据源是二维关系表，所以能够转化成定义 1 所述的信息系统并且属性制约的关联规则与粗糙集理论的决策规则的定义相对应，即在形成的规则的结果中，推出的结论只能有分类属性和对应值而条件属性只能作为前提，这样能够防止没用规则的出现。

4. 实验与讨论

4.1. 数据源

由于汽车在驾驶过程中，汽车的动力系统中，在某个档位，油门踏板、发动机的节气阀、发动机的转速和车速有时序的相关性，因此本文主要分析动力系统相关指令的相关性和时序性，因此我们采集动力系统 CAN 总线数据，采集设备是盘沣公司的 PFautoEcu-IV [7]，设备如图 1 所示，采集的部分数据集如图 2 所示，总共采集了 60 分钟的数据，约 40 万条数据。

4.2. 数据处理

首先对数据进行离散化处理，例如对 CAN 报文时间戳离散化处理过后，得到的断点集合为{0.0306, 0.0368, 0.0673, 0.0749, 0.1054, 0.1127, 0.1432, 0.1507, 0.1812, 0.19, 0.2192, 0.326, 0.3787, 0.469, 0.4849, 0.3407, 0.409, 0.333, …} (单位为 ms)。

Figure 1. Data acquisition equipment PFautoEcu-IV
图 1. 数据采集设备 PFautoEcu-IV

序号	CAN通道	传输	时间戳	状态	报文名称	帧ID	扩展帧	远程帧	DLC	数据		
1	Ox1	Receive	0.0306	status	0x301车速	0x301	FALSE	FALSE	8	0x[00	40	64
2	Ox1	Receive	0.0368	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 6F 00 00 00 00 00 00]		
3	Ox1	Receive	0.0673	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
4	Ox1	Receive	0.0749	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 6F 00 00 00 00 00 00]		
5	Ox1	Receive	0.1054	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
6	Ox1	Receive	0.1127	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 6F 00 00 00 00 00 00]		
7	Ox1	Receive	0.1432	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
8	Ox1	Receive	0.1507	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 74 00 00 00 00 00 00]		
9	Ox1	Receive	0.1812	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
10	Ox1	Receive	0.19	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 74 00 00 00 00 00 00]		
11	Ox1	Receive	0.2192	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
12	Ox1	Receive	0.2267	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 74 00 00 00 00 00 00]		
13	Ox1	Receive	0.257	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
14	Ox1	Receive	0.2649	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 6F 00 00 00 00 00 00]		
15	Ox1	Receive	0.2963	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
16	Ox1	Receive	0.3026	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 74 00 00 00 00 00 00]		
17	Ox1	Receive	0.333	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
18	Ox1	Receive	0.3407	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 74 00 00 00 00 00 00]		
19	Ox1	Receive	0.371	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
20	Ox1	Receive	0.3787	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 79 00 00 00 00 00 00]		
21	Ox1	Receive	0.409	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
22	Ox1	Receive	0.4175	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 6F 00 00 00 00 00 00]		
23	Ox1	Receive	0.4469	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
24	Ox1	Receive	0.4546	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 6F 00 00 00 00 00 00]		
25	Ox1	Receive	0.4849	status	0x301车速	0x301	FALSE	FALSE	8	0x[08 40 00 00 00 00 00 00]		
26	Ox1	Receive	0.4926	status	0x302转速	0x302	FALSE	FALSE	8	0x[08 6F 00 00 00 00 00 00]		

Figure 2. Partial data acquisition
图 2. 部分采集数据

试验得到的关联规则形如

$$\left|\text{时间戳差}\right| = [0.0060, 0.0080]$$

$$\text{关联规则} = \text{对应车速} -> \text{对应转速}$$

在上述规则中提取的是车速和转速的 CAN 报文，规则左边是条件属性，右边是结果。表示当 CAN 总线发送车速和转速的数据的时间戳差的绝对值在 0.0060~0.0080 之间那么说明此时车速和转速这 2 个属性就存在对应关系。这里形成规则的过程中，其中包含规则的记录有 230,000 条，置信度为 100%，支持度为 57.5%。

4.3. 结果分析

首先在测试数据源里添加异常数据，然后利用上面得到的关联规则对数据进行检测，得到试验部分结果如图 3 所示。从图中可以看出异常数据的时间戳差绝对值都没在范围内，正常数据都在范围内并且异常数据在车速确定的情况下，真实的转速值与添加的异常转速值不匹配，说明了达到了检测的目的。

车速 (km/h)	转速真实值 (rpm)	转速 (rpm)	车速时间戳 (ms)	转速时间戳 (ms)	时间戳差绝对值	是否在规则范围	数据情况
64	2159	2159	0.0306	0.0368	0.0062	是	正常
67	2332	2332	3.025	3.033	0.0082	是	正常
73	2410	3020	3.9855	4.0163	0.0308	否	异常
78	2620	2620	4.5083	4.5159	0.0076	是	正常
82	2919	2919	5.3802	5.3881	0.0079	是	正常
88	2939	3834	6.2537	6.2599	0.0026	否	异常
93	3233	3233	6.9358	6.9437	0.0079	是	正常
95	3432	3432	7.7698	7.7775	0.0077	是	正常
105	3678	3678	9.1735	9.1812	0.0077	是	正常
107	3476	3746	9.9311	9.938	0.0069	是	正常
122	4187	4187	11.5999	11.6076	0.0077	是	正常
128	4365	4365	12.7013	12.7076	0.0063	是	正常
139	4789	4789	14.3341	14.3416	0.0075	是	正常

Figure 3. Test result
图 3. 试验结果

5. 总结

本文通过粗糙集理论与关联规则挖掘算法的结合对车载网络动力系统的 CAN 报文进行了挖掘最后进行了试验，并对异常情况做出了检测，达到了预期的效果。但是还有不足的地方就是：这是针对数据的单维的挖掘和入侵检测，以后的研究方向可以把重心放到针对数据多维度的挖掘和入侵检测。

参考文献 (References)

[1] Miller, C. and Valasek, C. (2013) Adventures in Automotive Networks and Control Units. *DEF CON 21Hacking Conference*, Las Vegas, NV, 2-4 August 2013, 230-235.

[2] Miller, C. and Valasek, C. (2014) A Survey of Remote Automotive Attack Surfaces. *Black Hat*, USA, 2-7 August 2014, 1-94.

[3] Miller, C. and Valasek, C. (2015) Remote Exploitation of an Unaltered Passenger Vehicle. *Black Hat*, USA, 1-6 August 2015, 1-91.

[4] Murvay, P.-S. and Groza, B. (2014) Source Identification Using Signal Characteristics in Controller Area Networks. *IEEE Signal Processing Letters*, **21**, 395-399. https://doi.org/10.1109/LSP.2014.2304139

[5] Larson, U.E., Nilsson, D.K. and Jonsson, E. (2008) An Approach to Specification-Based Attack Detection for in-Vehicle Networks. 2008 *IEEE Intelligent Vehicles Symposium*, Eindhoven, 4-6 June 2008, 220-225. https://doi.org/10.1109/IVS.2008.4621263

[6] Muter, M., Groll, A. and Freiling, F.C. (2010) A Structured Approach to Anomaly Detection for in-Vehicle Networks. 2010 *Sixth International Conference on Information Assurance and Security*, Atlanta, GA, 23-25 August 2010, 92-98. https://doi.org/10.1109/ISIAS.2010.5604050

[7] 汽车电子与汽车 CAN 总线网络实验开发系统[EB/OL]. http://www.pfautocan.com/index.php?m=content&c=index&a=show&catid=8&id=22