(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107391598 B
(45)授权公告日 2021.01.26

(21)申请号 201710522318.X
(22)申请日 2017.06.30
(65)同一申请的已公布的文献号
申请公布号 CN 107391598 A
(43)申请公布日 2017.11.24
(73)专利权人 北京航空航天大学
地址 100191 北京市海淀区学院路37号
(72)发明人 李建欣 王婧仪 陈汉腾 李博 王嘉凯
(74)专利代理机构 北京中创阳光知识产权代理有限责任公司 11003
代理人 尹振启
(51)Int.Cl.
G06F 16/34(2019.01)
G06F 16/901(2019.01)

(56)对比文件
CN 106060018 A,2016.10.26
CN 106384048 A,2017.02.08
US 2016/0065599 A1,2016.03.03
周松松等.“基于威胁情报的恶意软件识别”.《信息安全等级保护技术大会入选论文》.2016,
Liao Xiaojing etc..“Acing the IOC Game:Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence”.《MARYLAND CYBERSECURITY CENTER》.2016,论文第10-25页.

审查员 李玥

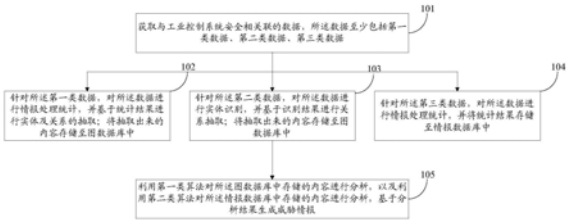
权利要求书2页 说明书7页 附图5页

(54)发明名称

一种威胁情报自动生成方法及系统

(57)摘要

本发明公开了一种威胁情报自动生成方法及系统,所述方法包括:获取与工业控制系统安全相关联的数据,所述数据至少包括第一类数据、第二类数据、第三类数据;针对所述第一类数据,对所述数据进行情报处理统计,并基于统计结果进行实体及关系的抽取;将抽取出来的内容存储至图数据库中;针对所述第二类数据,对所述数据进行实体识别,并基于识别结果进行关系抽取;将抽取出来的内容存储至图数据库中;针对所述第三类数据,对所述数据进行情报处理统计,并将统计结果存储至情报数据库中;利用第一类算法对所述图数据库中存储的内容进行分析,以及利用第二类算法对所述情报数据库中存储的内容进行分析,基于分析结果生成威胁情报。



CN 107391598 B

1.一种威胁情报自动生成方法,其特征在于,所述方法包括:

获取与工业控制系统安全相关联的数据,所述数据至少包括第一类数据、第二类数据、第三类数据;

所述获取与工业控制系统安全相关联的数据,包括:

利用蜜罐系统采集网络攻击流量数据,所述网络攻击流量数据属于所述第一类数据;

利用扫描系统采集工业控制设备分布数据及漏洞数据,所述工业控制设备分布数据及漏洞数据属于所述第一类数据;

采集来自互联网空间的数据,所述互联网空间的数据包括结构化数据、非结构化数据,其中,所述结构化数据属于所述第一类数据,所述非结构化数据属于所述第二类数据;

获取工业控制系统的开源安全威胁情报,所述开源安全威胁情报属于所述第三类数据;

针对所述蜜罐系统和所述扫描系统采集到的数据以及所述来自互联网空间的结构化数据,对所述数据进行情报处理统计,并基于统计结果进行实体及关系的抽取;将抽取出来的内容存储至图数据库中;

针对所述来自互联网空间的非结构化数据,利用机器学习方法对所述数据依次进行如下处理:文本规范化处理、文本分类处理、强相关文章提取处理;将提取出的强相关文章存储至强相关文章库中;对所述强相关文章库中的每篇文章逐句进行语法树分析以及正则匹配,并基于正则匹配结果进行正则过滤得到攻击指示器IOC item;基于所述语法树分析结果提取IOC item关系,并基于所述IOC item关系组建关系网;将所述关系网的关系信息存储至图数据库中;

针对所述第三类数据,对所述数据进行情报处理统计,并将统计结果存储至情报数据库中;

利用第一类算法对所述图数据库中存储的内容进行分析,以及利用第二类算法对所述情报数据库中存储的内容进行分析,基于分析结果生成威胁情报。

2.根据权利要求1所述的威胁情报自动生成方法,其特征在于,所述利用第一类算法对所述图数据库中存储的内容进行分析,以及利用第二类算法对所述情报数据库中存储的内容进行分析,基于分析结果生成威胁情报,包括:

针对所述图数据库中存储的内容,进行如下处理:聚类整合、数据融合、相似性分析、关联度分析;

针对所述情报数据库中存储的内容,进行如下处理:关键字查询、对比分析;

对处理后的内容进行威胁等级评定,基于威胁等级评定结果生成威胁情报。

3.一种威胁情报自动生成系统,其特征在于,所述系统包括:

数据采集模块,用于获取与工业控制系统安全相关联的数据,所述数据至少包括第一类数据、第二类数据、第三类数据;

所述数据采集模块,具体用于:

利用蜜罐系统采集网络攻击流量数据,所述网络攻击流量数据属于所述第一类数据;

利用扫描系统采集工业控制设备分布数据及漏洞数据,所述工业控制设备分布数据及漏洞数据属于所述第一类数据;

采集来自互联网空间的数据,所述互联网空间的数据包括结构化数据、非结构化数据,

其中,所述结构化数据属于所述第一类数据,所述非结构化数据属于所述第二类数据;

获取工业控制系统的开源安全威胁情报,所述开源安全威胁情报属于所述第三类数据;

数据处理模块,包括:第一处理子模块,用于针对所述蜜罐系统和所述扫描系统采集到的数据以及所述来自互联网空间的结构化数据,对所述数据进行情报处理统计,并基于统计结果进行实体及关系的抽取;将抽取出来的内容存储至图数据库中;第二处理子模块,用于针对所述来自互联网空间的非结构化数据,利用机器学习方法对所述数据依次进行如下处理:文本规范化处理、文本分类处理、强相关文章提取处理;将提取出的强相关文章存储至强相关文章库中;对所述强相关文章库中的每篇文章逐句进行语法树分析以及正则匹配,并基于正则匹配结果进行正则过滤得到IOC item;基于所述语法树分析结果提取IOC item关系,并基于所述IOC item关系组建关系网;将所述关系网的关系信息存储至图数据库中;针对所述第三类数据,对所述数据进行情报处理统计,并将统计结果存储至情报数据库中;

情报生成模块,用于利用第一类算法对所述图数据库中存储的内容进行分析,以及利用第二类算法对所述情报数据库中存储的内容进行分析,基于分析结果生成威胁情报。

4.根据权利要求3所述的威胁情报自动生成系统,其特征在于,所述情报生成模块,具体用于:

针对所述图数据库中存储的内容,进行如下处理:聚类整合、数据融合、相似性分析、关联度分析;

针对所述情报数据库中存储的内容,进行如下处理:关键字查询、对比分析;

对处理后的内容进行威胁等级评定,基于威胁等级评定结果生成威胁情报。

一种威胁情报自动生成方法及系统

技术领域

[0001] 本发明涉及工业控制系统安全技术领域，尤其涉及一种针对工业控制系统安全的威胁情报自动生成方法及系统。

背景技术

[0002] 工业控制系统(ICS，Industrial Control Systems)是由各种自动化控制组件和实时数据采集、监测的过程控制组件共同构成。ICS包括数据采集与监控系统(SCADA)、分布式控制系统(DCS)、可编程逻辑控制器(PLC)、远程终端(RTU)、智能电子设备(IED)，以及确保各组件通信的接口，被称为“系统中的系统”。ICS广泛应用于核设施、钢铁、有色、化工、石油石化、电力、天然气、先进制造等国家关键基础设施等领域，用于控制关键生产设备的运行。

[0003] 随着近年来“中国制造2025”、“互联网+”以及“工业4.0”计划的提出，在网络互连的大背景下，工业控制系统的互连已经成为不可避免的趋势。互连一方面可以提高生产力，提升创新能力，减少工业能源及资源消耗，助力产业模式转型升级，另一方面也会因为互联而诱发一系列网络安全问题，工业控制系统一直面临着来自内部和外部的各种恶意病毒的攻击。目前，工业控制系统遭受的网络攻击已经成为最严重的国家安全挑战之一。一些重点领域中的工业控制系统一旦遭到破坏，不仅会影响产业经济的持续发展，更会对国家安全造成巨大的损害。目前针对工业控制系统安全问题，仍主要采取传统安全防护措施，如防火墙、入侵检测、权限检测等。

[0004] 传统网络中威胁情报感知技术的作用对象单一、相互之间独立、粒度不够，给威胁情报的精确、完整、高效感知带来诸多不便。云计算虚拟化技术的出现极大地扩展了网络的规模，增加了信息系统的复杂性，给威胁情报感知带来了新的挑战。

发明内容

[0005] 为解决上述技术问题，本发明实施例提供了一种威胁情报自动生成方法及系统。

[0006] 本发明实施例提供的威胁情报自动生成方法，包括：

[0007] 获取与工业控制系统安全相关联的数据，所述数据至少包括第一类数据、第二类数据、第三类数据；

[0008] 针对所述第一类数据，对所述数据进行情报处理统计，并基于统计结果进行实体及关系的抽取；将抽取出来的内容存储至图数据库中；

[0009] 针对所述第二类数据，对所述数据进行实体识别，并基于识别结果进行关系抽取；将抽取出来的内容存储至图数据库中；

[0010] 针对所述第三类数据，对所述数据进行情报处理统计，并将统计结果存储至情报数据库中；

[0011] 利用第一类算法对所述图数据库中存储的内容进行分析，以及利用第二类算法对所述情报数据库中存储的内容进行分析，基于分析结果生成威胁情报。

[0012] 本发明实施例中，所述获取与工业控制系统安全相关联的数据，包括：

[0013] 利用蜜罐系统采集网络攻击流量数据，所述网络攻击流量数据属于所述第一类数据；

[0014] 利用扫描系统采集工业控制设备分布数据及漏洞数据，所述工业控制设备分布数据及漏洞数据属于所述第一类数据；

[0015] 采集来自互联网空间的数据，所述互联网空间的数据包括结构化数据、非结构化数据，其中，所述结构化数据属于所述第一类数据，所述非结构化数据属于所述第二类数据；

[0016] 获取工业控制系统的开源安全威胁情报，所述开源安全威胁情报属于所述第三类数据。

[0017] 本发明实施例中，所述针对所述第一类数据，对所述数据进行情报处理统计，并基于统计结果进行实体及关系的抽取；将抽取出来的内容存储至图数据库中，包括：

[0018] 针对所述蜜罐系统和所述扫描系统采集到的数据以及所述来自互联网空间的结构化数据，对所述数据进行情报处理统计，并基于统计结果进行实体及关系的抽取；将抽取出来的内容存储至图数据库中。

[0019] 本发明实施例中，所述针对所述第二类数据，对所述数据进行实体识别，并基于识别结果进行关系抽取；将抽取出来的内容存储至图数据库中，包括：

[0020] 针对所述来自互联网空间的非结构化数据，利用机器学习方法对所述数据依次进行如下处理：文本规范化处理、文本分类处理、强相关文章提取处理；

[0021] 将提取出的强相关文章存储至强相关文章库中；

[0022] 对所述强相关文章库中的每篇文章逐句进行语法树分析以及正则匹配，并基于正则匹配结果进行正则过滤得到攻击指示器(IOC item)；

[0023] 基于所述语法树分析结果提取IOC item关系，并基于所述IOC item关系组建关系网；

[0024] 将所述关系网的关系信息存储至图数据库中。

[0025] 本发明实施例中，所述利用第一类算法对所述图数据库中存储的内容进行分析，以及利用第二类算法对所述情报数据库中存储的内容进行分析，基于分析结果生成威胁情报，包括：

[0026] 针对所述图数据库中存储的内容，进行如下处理：聚类整合、数据融合、相似性分析、关联度分析；

[0027] 针对所述情报数据库中存储的内容，进行如下处理：关键字查询、对比分析；

[0028] 对处理后的内容进行威胁等级评定，基于威胁等级评定结果生成威胁情报。

[0029] 本发明实施例提供的威胁情报自动生成系统，包括：

[0030] 数据采集模块，用于获取与工业控制系统安全相关联的数据，所述数据至少包括第一类数据、第二类数据、第三类数据；

[0031] 数据处理模块，用于针对所述第一类数据，对所述数据进行情报处理统计，并基于统计结果进行实体及关系的抽取；将抽取出来的内容存储至图数据库中；针对所述第二类数据，对所述数据进行实体识别，并基于识别结果进行关系抽取；将抽取出来的内容存储至图数据库中；针对所述第三类数据，对所述数据进行情报处理统计，并将统计结果存储至情

报数据库中；

[0032] 情报生成模块，用于利用第一类算法对所述图数据库中存储的内容进行分析，以及利用第二类算法对所述情报数据库中存储的内容进行分析，基于分析结果生成威胁情报。

[0033] 本发明实施例中，所述数据采集模块，具体用于：

[0034] 利用蜜罐系统采集网络攻击流量数据，所述网络攻击流量数据属于所述第一类数据；

[0035] 利用扫描系统采集工业控制设备分布数据及漏洞数据，所述工业控制设备分布数据及漏洞数据属于所述第一类数据；

[0036] 采集来自互联网空间的数据，所述互联网空间的数据包括结构化数据、非结构化数据，其中，所述结构化数据属于所述第一类数据，所述非结构化数据属于所述第二类数据；

[0037] 获取工业控制系统的开源安全威胁情报，所述开源安全威胁情报属于所述第三类数据。

[0038] 本发明实施例中，所述数据处理模块，包括：第一处理子模块，用于针对所述蜜罐系统和所述扫描系统采集到的数据以及所述来自互联网空间的结构化数据，对所述数据进行情报处理统计，并基于统计结果进行实体及关系的抽取；将抽取出来的内容存储至图数据库中。

[0039] 本发明实施例中，所述数据处理模块，包括：第二处理子模块，用于针对所述来自互联网空间的非结构化数据，利用机器学习方法对所述数据依次进行如下处理：文本规范化处理、文本分类处理、强相关文章提取处理；将提取出的强相关文章存储至强相关文章库中；对所述强相关文章库中的每篇文章逐句进行语法树分析以及正则匹配，并基于正则匹配结果进行正则过滤得到IOC item；基于所述语法树分析结果提取IOC item关系，并基于所述IOC item关系组建关系网；将所述关系网的关系信息存储至图数据库中。

[0040] 本发明实施例中，所述情报生成模块，具体用于：

[0041] 针对所述图数据库中存储的内容，进行如下处理：聚类整合、数据融合、相似性分析、关联度分析；

[0042] 针对所述情报数据库中存储的内容，进行如下处理：关键字查询、对比分析；

[0043] 对处理后的内容进行威胁等级评定，基于威胁等级评定结果生成威胁情报。

[0044] 本发明实施例还提供一种计算机存储介质，其上存储有计算机可执行指令，其特征在于，该计算机可执行指令被处理器执行时实现上述任意所述的威胁情报自动生成方法。

[0045] 本发明实施例提供的计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机可执行指令，其特征在于，所述处理器执行所述计算机可执行指令时实现上述任意所述的威胁情报自动生成方法。

[0046] 本发明实施例的技术方案中，通过多种方式自动采集互联网空间中工业控制系统相关信息，运用机器学习及图数据库等技术手段分析处理数据，最终输出格式化的威胁情报，该威胁情报可提供关于现存的、或者是即将出现的针对工业控制系统的威胁或危险的信息，为相关部门或企业响应相关威胁或危险提供决策支撑。

附图说明

[0047] 图1为本发明实施例的威胁情报自动生成方法的流程示意图；

[0048] 图2为本发明实施例的威胁情报自动生成系统的结构组成示意图一；

[0049] 图3为本发明实施例的威胁情报自动生成系统的结构组成示意图二；

[0050] 图4为本发明实施例的非结构化数据的处理流程图；

[0051] 图5为本发明实施例的计算机设备的结构组成示意图。

具体实施方式

[0052] 为了能够更加详尽地了解本发明实施例的特点与技术内容，下面结合附图对本发明实施例的实现进行详细阐述，所附附图仅供参考说明之用，并非用来限定本发明实施例。

[0053] 图1为本发明实施例的威胁情报自动生成方法的流程示意图，如图1所示，所述威胁情报自动生成方法包括以下步骤：

[0054] 步骤101：获取与工业控制系统安全相关联的数据，所述数据至少包括第一类数据、第二类数据、第三类数据。

[0055] 具体地，利用蜜罐系统采集网络攻击流量数据，所述网络攻击流量数据属于所述第一类数据；

[0056] 利用扫描系统采集工业控制设备分布数据及漏洞数据，所述工业控制设备分布数据及漏洞数据属于所述第一类数据；

[0057] 采集来自互联网空间的数据，所述互联网空间的数据包括结构化数据、非结构化数据，其中，所述结构化数据属于所述第一类数据，所述非结构化数据属于所述第二类数据；

[0058] 获取工业控制系统的开源安全威胁情报，所述开源安全威胁情报属于所述第三类数据。

[0059] 步骤102：针对所述第一类数据，对所述数据进行情报处理统计，并基于统计结果进行实体及关系的抽取；将抽取出来的内容存储至图数据库中。

[0060] 具体地，针对所述蜜罐系统和所述扫描系统采集到的数据以及所述来自互联网空间的结构化数据，对所述数据进行情报处理统计，并基于统计结果进行实体及关系的抽取；将抽取出来的内容存储至图数据库中。

[0061] 步骤103：针对所述第二类数据，对所述数据进行实体识别，并基于识别结果进行关系抽取；将抽取出来的内容存储至图数据库中。

[0062] 具体地，针对所述来自互联网空间的非结构化数据，利用机器学习方法对所述数据依次进行如下处理：文本规范化处理、文本分类处理、强相关文章提取处理；

[0063] 将提取出的强相关文章存储至强相关文章库中；

[0064] 对所述强相关文章库中的每篇文章逐句进行语法树分析以及正则匹配，并基于正则匹配结果进行正则过滤得到IOC item；

[0065] 基于所述语法树分析结果提取IOC item关系，并基于所述IOC item关系组建关系网；

[0066] 将所述关系网的关系信息存储至图数据库中。

[0067] 步骤104：针对所述第三类数据，对所述数据进行情报处理统计，并将统计结果存

储至情报数据库中。

[0068] 步骤105:利用第一类算法对所述图数据库中存储的内容进行分析,以及利用第二类算法对所述情报数据库中存储的内容进行分析,基于分析结果生成威胁情报。

[0069] 具体地,针对所述图数据库中存储的内容,进行如下处理:聚类整合、数据融合、相似性分析、关联度分析;

[0070] 针对所述情报数据库中存储的内容,进行如下处理:关键字查询、对比分析;

[0071] 对处理后的内容进行威胁等级评定,基于威胁等级评定结果生成威胁情报。

[0072] 上述方案中,步骤102、步骤103以及步骤104之间不限定执行先后顺序。

[0073] 图2为本发明实施例的威胁情报自动生成系统的结构组成示意图一,如图2所示,所述威胁情报自动生成系统包括:

[0074] 数据采集模块201,用于获取与工业控制系统安全相关联的数据,所述数据至少包括第一类数据、第二类数据、第三类数据;

[0075] 数据处理模块202,用于针对所述第一类数据,对所述数据进行情报处理统计,并基于统计结果进行实体及关系的抽取;将抽取出来的内容存储至图数据库中;针对所述第二类数据,对所述数据进行实体识别,并基于识别结果进行关系抽取;将抽取出来的内容存储至图数据库中;针对所述第三类数据,对所述数据进行情报处理统计,并将统计结果存储至情报数据库中;

[0076] 情报生成模块203,用于利用第一类算法对所述图数据库中存储的内容进行分析,以及利用第二类算法对所述情报数据库中存储的内容进行分析,基于分析结果生成威胁情报。

[0077] 本发明实施例中,所述数据采集模块201,具体用于:

[0078] 利用蜜罐系统采集网络攻击流量数据,所述网络攻击流量数据属于所述第一类数据;

[0079] 利用扫描系统采集工业控制设备分布数据及漏洞数据,所述工业控制设备分布数据及漏洞数据属于所述第一类数据;

[0080] 采集来自互联网空间的数据,所述互联网空间的数据包括结构化数据、非结构化数据,其中,所述结构化数据属于所述第一类数据,所述非结构化数据属于所述第二类数据;

[0081] 获取工业控制系统的开源安全威胁情报,所述开源安全威胁情报属于所述第三类数据。

[0082] 本发明实施例中,所述数据处理模块202,包括:第一处理子模块2021,用于针对所述蜜罐系统和所述扫描系统采集到的数据以及所述来自互联网空间的结构化数据,对所述数据进行情报处理统计,并基于统计结果进行实体及关系的抽取;将抽取出来的内容存储至图数据库中。

[0083] 本发明实施例中,所述数据处理模块202,包括:第二处理子模块2022,用于针对所述来自互联网空间的非结构化数据,利用机器学习方法对所述数据依次进行如下处理:文本规范化处理、文本分类处理、强相关文章提取处理;将提取出的强相关文章存储至强相关文章库中;对所述强相关文章库中的每篇文章逐句进行语法树分析以及正则匹配,并基于正则匹配结果进行正则过滤得到IOC item;基于所述语法树分析结果提取IOC item关系,

并基于所述IOC item关系组建关系网;将所述关系网的关系信息存储至图数据库中。

[0084] 本发明实施例中,所述情报生成模块203,具体用于:

[0085] 针对所述图数据库中存储的内容,进行如下处理:聚类整合、数据融合、相似性分析、关联度分析;

[0086] 针对所述情报数据库中存储的内容,进行如下处理:关键字查询、对比分析;

[0087] 对处理后的内容进行威胁等级评定,基于威胁等级评定结果生成威胁情报。

[0088] 图3为本发明实施例的威胁情报自动生成系统的结构组成示意图二,如图3所示,所述威胁情报自动生成系统包括:数据采集子系统、数据处理子系统、情报生成子系统。下面分别对各个子系统进行如下描述:

[0089] (1)数据采集子系统

[0090] 数据采集系统负责自动获取工控安全相关数据,数据主要有4个来源:蜜罐系统获取的网络攻击流量数据、扫描系统获取的全球工控设备分布及相关漏洞数据、来自互联网空间的数据及开源的工业控制系统安全威胁情报。其中来自互联网空间的数据分为两类:一类是来自漏洞库、补丁库等的结构化数据;另一类是来自论文、新闻、微博、微信、技术论坛等的非结构化文本数据。

[0091] (2)数据处理子系统

[0092] 数据处理子系统是使用图数据库及机器学习等技术构成的综合系统,收集到来自蜜罐系统、扫描系统及来自互联网空间的数据后存入图数据库。

[0093] 针对蜜罐系统和扫描系统采集到的数据,进行情报统计处理后,根据预定义的实体及关系,将数据抽取并存入图数据库。针对来自互联网空间的结构化数据,同样进行信息抽取后存入图数据库。

[0094] 针对来自互联网空间的非结构化数据,如图4所示,利用机器学习的技术,对非结构化的文本数据文本规范化、分类,提取出强相关文章,存入强相关文章库。对库中的每篇文章逐句进行语法树分析,并进行正则匹配,经过正则过滤后得到IOC item。根据分析结果提取IOC item关系,组建关系网,并将关系信息存入图数据库。

[0095] 针对收集到的开源情报,进行情报处理统计后存入情报数据库。

[0096] (3)情报生成子系统

[0097] 情报生成子系统,在图数据库中利用相关图算法,对数据进行聚类整合、数据融合,并进行相似性分析及关联度分析。同时,对收集到的开源威胁情报,在图数据库中进行相关关键字查询,根据已有的数据,进行对比分析,增加情报富含度及可信度,最后经过威胁等级评定,输出威胁情报。

[0098] 本发明实施例上述系统如果以软件功能模块的形式实现并作为独立的产品销售或使用时,也可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明实施例的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机、服务器、或者网络设备等)执行本发明各个实施例所述方法的全部或部分。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read Only Memory)、磁碟或者光盘等各种可以存储程序代码的介质。这样,本发明实施例不限制于任何特定的硬件和软件结合。

[0099] 相应地,本发明实施例还提供一种计算机存储介质,其中存储有计算机可执行指

令，该计算机可执行指令被处理器执行时实现本发明实施例的上述威胁情报自动生成方法。

[0100] 图5为本发明实施例的计算机设备的结构组成示意图，如图5所示，所述计算机设备包括存储器501、处理器502及存储在存储器501上并可在处理器502上运行的计算机可执行指令，所述处理器502执行所述计算机可执行指令时实现如下方法步骤：

[0101] 获取与工业控制系统安全相关联的数据，所述数据至少包括第一类数据、第二类数据、第三类数据；

[0102] 针对所述第一类数据，对所述数据进行情报处理统计，并基于统计结果进行实体及关系的抽取；将抽取出来的内容存储至图数据库中；

[0103] 针对所述第二类数据，对所述数据进行实体识别，并基于识别结果进行关系抽取；将抽取出来的内容存储至图数据库中；

[0104] 针对所述第三类数据，对所述数据进行情报处理统计，并将统计结果存储至情报数据库中；

[0105] 利用第一类算法对所述图数据库中存储的内容进行分析，以及利用第二类算法对所述情报数据库中存储的内容进行分析，基于分析结果生成威胁情报。

[0106] 以上涉及计算机设备的描述，与上述方法描述是类似的，同方法的有益效果描述，不做赘述。

[0107] 本发明实施例所记载的技术方案之间，在不冲突的情况下，可以任意组合。

[0108] 在本发明所提供的几个实施例中，应该理解到，所揭露的方法和智能设备，可以通过其它的方式实现。以上所描述的设备实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，如：多个单元或组件可以结合，或可以集成到另一个系统，或一些特征可以忽略，或不执行。另外，所显示或讨论的各组成部分相互之间的耦合、或直接耦合、或通信连接可以是通过一些接口，设备或单元的间接耦合或通信连接，可以是电性的、机械的或其它形式的。

[0109] 上述作为分离部件说明的单元可以是、或也可以不是物理上分开的，作为单元显示的部件可以是、或也可以不是物理单元，即可以位于一个地方，也可以分布到多个网络单元上；可以根据实际的需要选择其中的部分或全部单元来实现本实施例方案的目的。

[0110] 另外，在本发明各实施例中的各功能单元可以全部集成在一个第二处理单元中，也可以是各单元分别单独作为一个单元，也可以两个或两个以上单元集成在一个单元中；上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能单元的形式实现。

[0111] 以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本发明的保护范围之内。

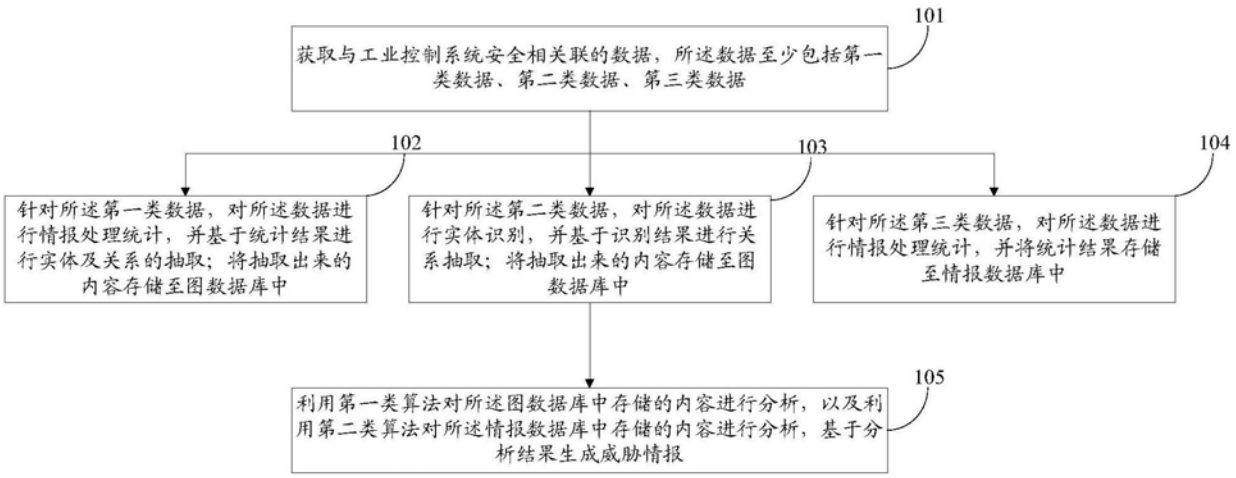

101
获取与工业控制系统安全相关联的数据，所述数据至少包括第一类数据、第二类数据、第三类数据
102
103
104
针对所述第一类数据，对所述数据进行情报处理统计，并基于统计结果进行实体及关系的抽取；将抽取出来的内容存储至图数据库中
针对所述第二类数据，对所述数据进行实体识别，并基于识别结果进行关系抽取；将抽取出来的内容存储至图数据库中
针对所述第三类数据，对所述数据进行情报处理统计，并将统计结果存储至情报数据库中
105
利用第一类算法对所述图数据库中存储的内容进行分析，以及利用第二类算法对所述情报数据库中存储的内容进行分析，基于分析结果生成威胁情报

图1

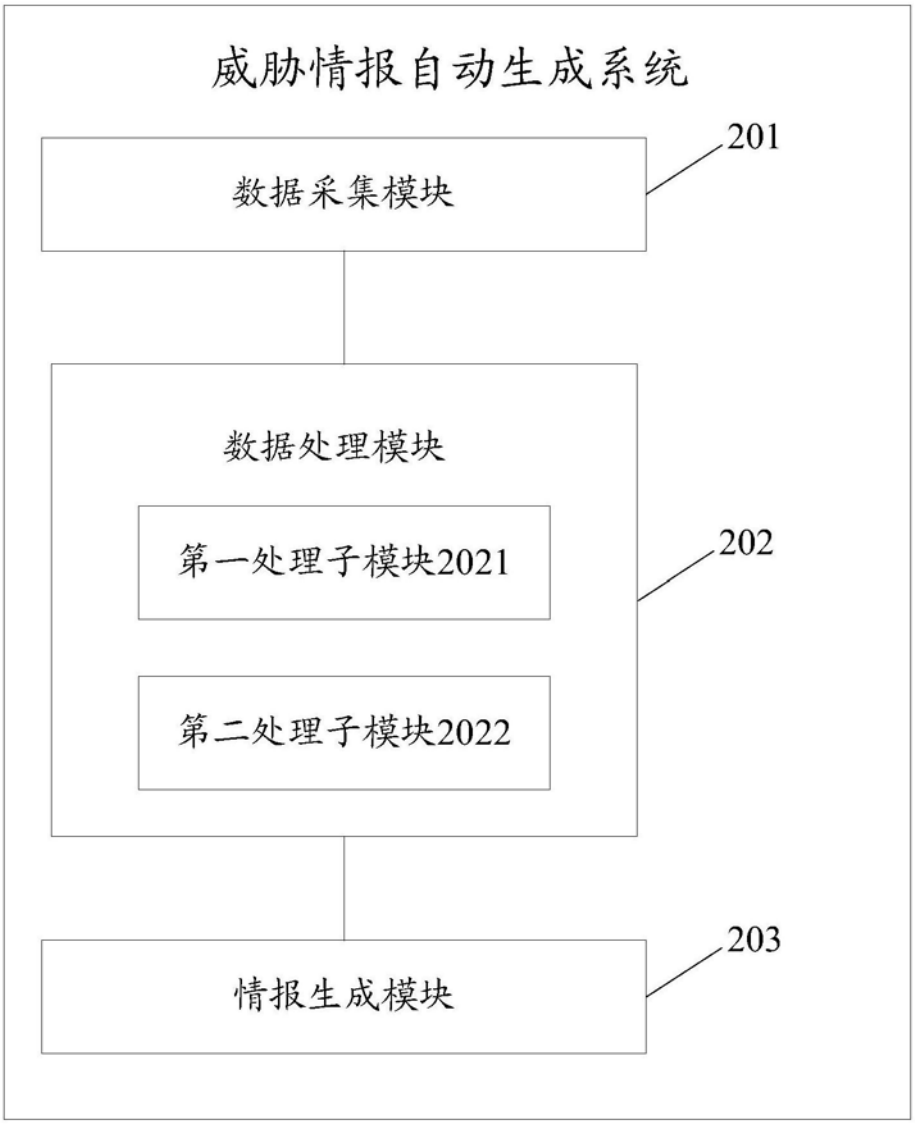

威胁情报自动生成系统
201
数据采集模块
数据处理模块
第一处理子模块2021
第二处理子模块2022
202
203
情报生成模块

图2

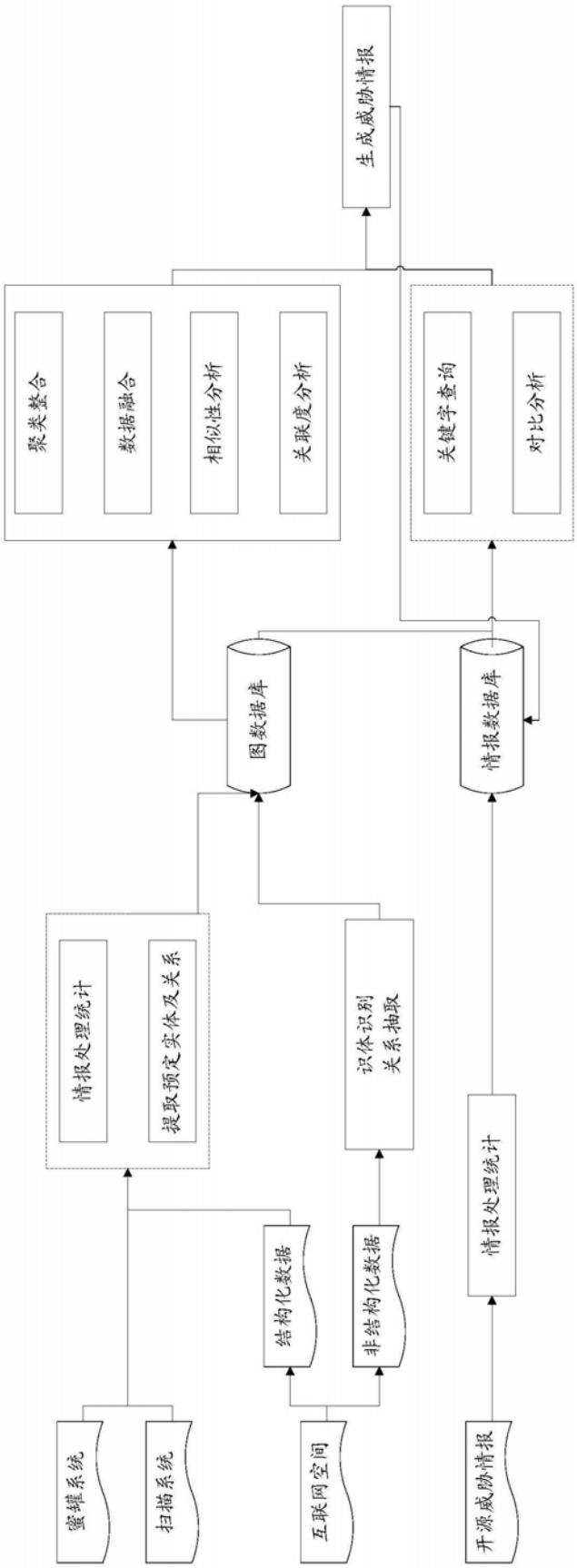

生成威胁情报
聚类整合
数据融合
相似性分析
关联度分析
关键字查询
对比分析
图数据库
情报数据库
情报处理统计
提取预定实体及关系
识体识别
关系抽取
结构化数据
非结构化数据
情报处理统计
蜜罐系统
扫描系统
互联网空间
开源威胁情报

图3

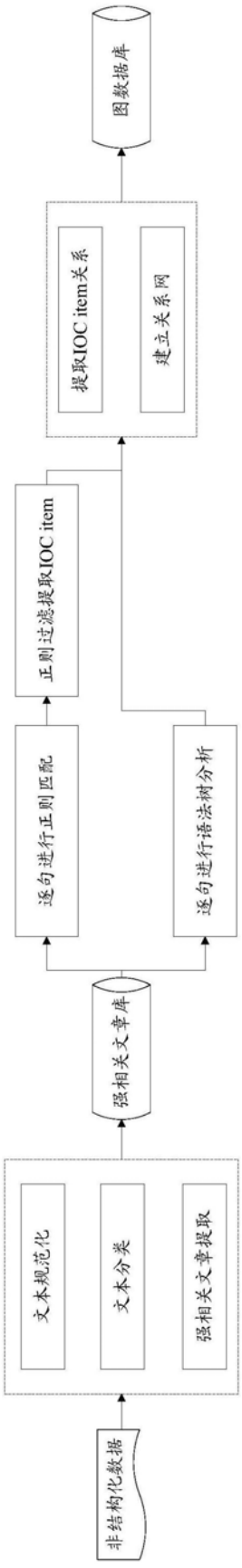

非结构化数据
文本规范化
文本分类
强相关文章提取
强相关文章库
逐句进行正则匹配
逐句进行语法树分析
正则过滤提取IOC item
提取IOC item关系
建立关系网
图数据库

图4

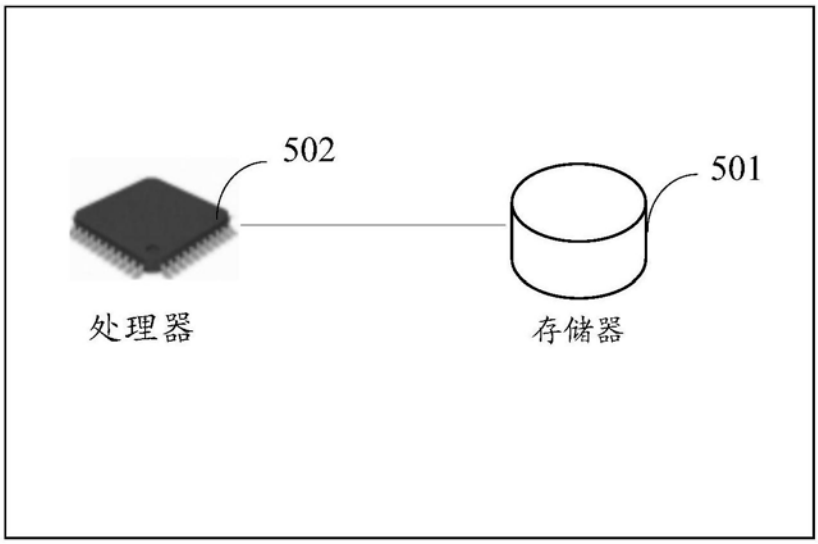


图5