

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2021 年 12 月 9 日 (09.12.2021)



(10) 国际公布号
WO 2021/243740 A1

- (51) 国际专利分类号:
G06F 21/54 (2013.01)
- (21) 国际申请号: PCT/CN2020/095248
- (22) 国际申请日: 2020 年 6 月 10 日 (10.06.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202010489633.9 2020年6月2日 (02.06.2020) CN
- (71) 申请人: 浙江大学(ZHEJIANG UNIVERSITY) [CN/CN]; 中国浙江省杭州市余杭塘路 866 号, Zhejiang 310058 (CN)。
- (72) 发明人: 吴斯韦(WU, Siwei); 中国浙江省杭州市余杭塘路866号, Zhejiang 310058 (CN)。 周亚金

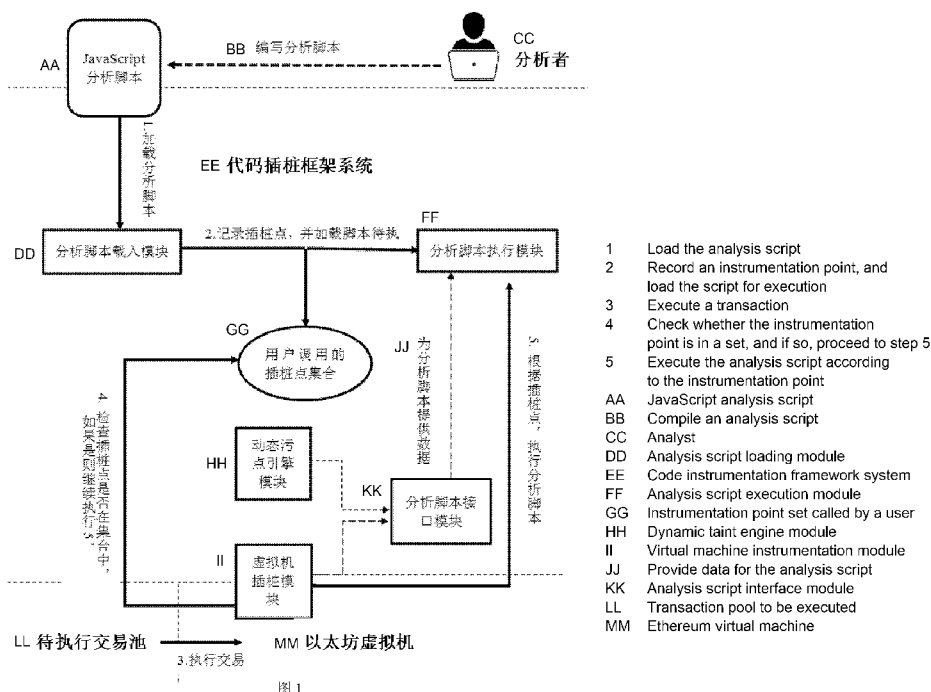
(ZHOU, Yajin); 中国浙江省杭州市余杭塘路866号, Zhejiang 310058 (CN)。 吴磊(WU, Lei); 中国浙江省杭州市余杭塘路866号, Zhejiang 310058 (CN)。 任奎(REN, Kui); 中国浙江省杭州市余杭塘路866号, Zhejiang 310058 (CN)。

(74) 代理人: 杭州求是专利事务所有限公司(HANGZHOU QIUSHI PATENT OFFICE CO., LTD.); 中国浙江省杭州市玉古路 147 号黄鸿年科技综合大楼 215 室贾玉霞, Zhejiang 310013 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) **Title:** CODE INSTRUMENTATION FRAMEWORK SYSTEM BASED ON ETHEREUM VIRTUAL MACHINE

(54) 发明名称: 一种基于以太坊虚拟机的代码插桩框架系统



(57) **Abstract:** A code instrumentation framework system based on an Ethereum virtual machine. The system comprises an analysis script loading module, an analysis script execution module, an Ethereum virtual machine instrumentation module, a dynamic taint engine module and an analysis script interface module, wherein the Ethereum virtual machine instrumentation module inserts three types of instrumentation points into an Ethereum virtual machine to interrupt the execution of the Ethereum virtual machine, and also provides, for the analysis script interface module, various data supports during the execution process of the virtual machine; the analysis

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

script loading module is used for loading an analysis script and recording an instrumentation point in which a user is interested; the dynamic taint engine module is used for providing, for the analysis script interface module, various data supports related to data flow analysis; and the analysis script interface module provides an interface for the analysis script execution module so as to view various pieces of data during the execution process of the virtual machine and various pieces of data related to data flow analysis. By providing various interfaces, an analyst is assisted in reducing development costs, and operation performance loss is greatly reduced.

(57) 摘要: 一种基于以太坊虚拟机的代码插桩框架系统, 包括分析脚本载入模块、分析脚本执行模块、以太坊虚拟机插桩模块、动态污点引擎模块、分析脚本接口模块; 以太坊虚拟机插桩模块一方面在以太坊虚拟机中插入三类插桩点, 用于打断以太坊虚拟机的执行, 另一方面为分析脚本接口模块提供虚拟机在执行过程中的各类数据支持; 分析脚本载入模块用于载入分析脚本并记录用户关注的插桩点; 动态污点引擎模块用于为分析脚本接口模块提供数据流分析相关的各类数据支持; 分析脚本接口模块为分析脚本执行模块提供接口查看虚拟机执行过程中各类数据和数据流分析相关的各类数据。通过提供丰富的接口, 帮助分析人员减小开发成本, 大大减少运行性能损耗。

一种基于以太坊虚拟机的代码插桩框架系统

技术领域

本发明涉及程序分析信息技术领域，具体涉及一种基于以太坊虚拟机的代码插桩框架系统。

背景技术

近年来，随着比特币系统的平稳运行，区块链技术受到世界范围内广泛的关注，而以太坊的诞生又是一个里程碑意义的事件。以太坊作为市值排名第二的区块链平台，最重要的突破和特色就是配备了图灵完备的以太坊虚拟机，可以支持智能合约的执行。以太坊原生的以太币以及在以太坊上发行的众多数字货币已经具备了相当的金融属性。以太坊网络由运行着以太坊客户端软件的节点组成，主要包括 Go-Ethereum（简称 Geth）和 Parity 等。

由于区块链技术，以太坊平台，智能合约都是新生事物，又因其具备的金融属性，近些年吸引了许多的恶意攻击者在以太坊平台上兴风作浪，严重地影响了以太坊的生态，比如臭名昭著的重入攻击，曾对以太坊平台造成巨大的影响。

程序插桩技术，是在保证被测程序原有逻辑完整性的基础上在程序中插入一些“探针”，本质上就是进行信息采集的代码段，通过“探针”的执行抛出程序运行的特征数据，通过对这些数据的分析，可以获取程序的控制流和数据流信息，进而得到逻辑覆盖等动态信息。程序插桩被广泛用于传统程序分析中。

随着以太坊平台上各类攻击漏洞层出不穷，结合程序插桩技术检测恶意行为的需求也愈发急迫，而近年来已经出现了一些基于插桩技术的攻击检测方法，不过由于没有成熟的插桩框架出现，目前以太坊与插桩技术的结合仍然是通过直接对以太坊虚拟机插入“探针”的方式，这需要分析人员对以太坊虚拟机的代码逻辑非常了解，并且针对不同类型攻击的检测会需要不同的插桩方式，不仅需要分析人员投入很高的开发成本，而且缺乏可扩展性。

Geth 内嵌的 trace 功能提供了一个朴素的插桩框架，它在执行时会在每一个操作码(opcode)执行之前跳转到插桩代码的执行环境，而不管用户是否在意该种操作码(opcode)，此种设计会带来额外的上下文切换开销，无法应对大规模的实验。并且它不具备动态污点引擎，分析人员在使用时仍然需要付出很高的开发成本来追踪程序的数据流。

发明内容

为了解决现有技术中存在的上述缺陷，本发明的公开了一种基于以太坊虚拟机的代码插桩框架系统，该系统可以帮助用户以最低的开发成本深入到以太坊虚拟机层面来理解和检测

恶意行为。具体技术方案如下：

一种基于以太坊虚拟机的代码插桩框架系统，其特征在于，该系统包括分析脚本载入模块、分析脚本执行模块、以太坊虚拟机插桩模块、动态污点引擎模块和分析脚本接口模块；

所述的以太坊虚拟机插桩模块一方面在以太坊虚拟机中插入三类插桩点，用于打断以太坊虚拟机的执行，另一方面为所述的分析脚本接口模块提供以太坊虚拟机在执行过程中的各类数据支持；所述的三类插桩点包括常规交易类插桩点、智能合约类插桩点和操作码类插桩点；所述的常规交易指的是由非智能合约账户发起的交易；

所述的分析脚本载入模块用于载入分析脚本并记录用户在乎的插桩点；

所述的分析脚本执行模块用于执行分析脚本；

所述的动态污点引擎模块用于为所述的分析脚本接口模块提供数据流分析相关的各类数据支持；

所述的分析脚本接口模块为所述的分析脚本执行模块提供接口，查看以太坊虚拟机在执行过程中的各类数据和数据流分析相关的各类数据。

进一步地，所述的常规交易类插桩点包含常规交易的开始和常规交易的结束。

进一步地，所述的智能合约插桩点包含执行一个智能合约的开始和结束。

进一步地，所述的操作码类插桩点包含操作码执行之前和操作码执行之后。

进一步地，所述的虚拟机在执行过程中的各类数据包括以太坊虚拟机中的栈、内存，交易输入字段、所涉及智能合约的状态变量和交易返回字段、当前交易的交易哈希和所处区块号；

其中，所述的栈、内存指的是以太坊虚拟机中的数据存储结构。

进一步地，所述的数据流分析相关的各类数据指的是动态污点引擎模块自主维护的用于存储和传播污点的影子栈、影子内存、影子交易输入字段、所涉及智能合约的影子状态变量和影子交易返回字段。

本发明的有益效果如下：

本发明的代码插桩框架系统可以帮助分析人员深入到以太坊虚拟机层面来理解和检测恶意行为，通过提供灵活的插桩接口，在完成基于以太坊虚拟机插桩的交易分析任务中可以节省用户大量的开发成本；通过提供查看以太坊虚拟机在执行过程中的各类数据的接口，节省了用户理解以太坊虚拟机底层逻辑的开发成本；通过配备动态污点引擎提供的查看数据流相关的各类数据的接口，节省了用户记录数据流的开发成本；通过提前加载用户脚本并记录用户在乎的插桩点，可以做到在执行时按需调用用户的分析脚本，避免不必要的上下文切换而带来的性能损耗。

附图说明

图 1 为本发明的基于以太坊虚拟机的代码插桩框架系统的示意图。

具体实施方式

下面根据附图和优选实施例详细描述本发明，本发明的目的和效果将变得更加明白，应当理解，此处所描述的具体实施例仅仅用于解释本发明，并不用于限定本发明。

如图 1 所示，本发明的基于以太坊的代码插桩框架系统，包括：

(1) 以太坊虚拟机插桩模块，该模块一方面在以太坊虚拟机中插入三类插桩点，用于打断以太坊虚拟机的执行，另一方面为分析脚本接口模块提供以太坊虚拟机在执行过程中的各类数据支持；所述的三类插桩点包括：

①常规交易类插桩点：常规交易的开始和常规交易的结束。

②智能合约类插桩点：执行一个智能合约的开始和结束。

③操作码类插桩点：操作码执行之前和操作码执行之后。

所述的常规交易指的是由非智能合约账户发起的交易。

所述的虚拟机在执行过程中的各类数据包括以太坊虚拟机中的栈、内存、交易输入字段，所涉及智能合约的状态变量和交易返回字段、当前交易的交易哈希和所处区块号；

所述的栈、内存指的是以太坊虚拟机中的数据存储结构。

(2) 分析脚本载入模块，该模块用于载入分析脚本并记录用户在意的插桩点，目的是为了减少不必要的上下文切换而带来的性能损耗；所述的上下文切换指的是从以太坊虚拟机的执行上下文切换到分析脚本引擎的执行上下文。

(3) 分析脚本执行模块，该模块用于执行分析脚本。

(4) 动态污点引擎模块，该模块用于为分析脚本接口模块提供数据流分析相关的各类数据支持；所述的数据流分析相关的各类数据指的是动态污点引擎模块自主维护的用于存储和传播污点的影子栈、影子内存、影子交易输入字段、所涉及智能合约的影子状态变量和影子交易返回字段。

(5) 分析脚本接口模块，该模块用于为分析脚本执行模块提供接口，查看以太坊虚拟机在执行过程中的各类数据和数据流分析相关的各类数据。

如图 1 所示，本发明的基于以太坊虚拟机的代码插桩框架系统部署在 Geth 内嵌的以太坊虚拟机中，其中分析脚本执行模块采用的是 go-duktape 提供的 JavaScript 执行引擎，为用户提供的开发脚本语言是 JavaScript。

本领域普通技术人员可以理解，以上所述仅为发明的优选实例而已，并不用于限制发明，尽管参照前述实例对发明进行了详细的说明，对于本领域的技术人员来说，其依然可以对前

述各实例记载的技术方案进行修改，或者对其中部分技术特征进行等同替换。凡在发明的精神和原则之内，所做的修改、等同替换等均应包含在发明的保护范围之内。

权利要求书

1. 一种基于以太坊虚拟机的代码插桩框架系统，其特征在于，该系统包括分析脚本载入模块、分析脚本执行模块、以太坊虚拟机插桩模块、动态污点引擎模块和分析脚本接口模块；

所述的以太坊虚拟机插桩模块一方面在以太坊虚拟机中插入三类插桩点，用于打断以太坊虚拟机的执行，另一方面为所述的分析脚本接口模块提供以太坊虚拟机在执行过程中的各类数据支持；所述的三类插桩点包括常规交易类插桩点、智能合约类插桩点和操作码类插桩点；所述的常规交易指的是由非智能合约账户发起的交易。

所述的分析脚本载入模块用于载入分析脚本并记录用户在意的插桩点；

所述的分析脚本执行模块用于执行分析脚本；

所述的动态污点引擎模块用于为所述的分析脚本接口模块提供数据流分析相关的各类数据支持；

所述的分析脚本接口模块为所述的分析脚本执行模块提供接口，查看以太坊虚拟机在执行过程中的各类数据和数据流分析相关的各类数据。

2. 根据权利要求1所述的基于以太坊虚拟机的代码插桩框架系统，其特征在于，所述的常规交易类插桩点包含常规交易的开始和常规交易的结束。

3. 根据权利要求1所述的基于以太坊虚拟机的代码插桩框架系统，其特征在于，所述的智能合约插桩点包含执行一个智能合约的开始和结束。

4. 根据权利要求1所述的基于以太坊虚拟机的代码插桩框架系统，其特征在于，所述的操作码类插桩点包含操作码执行之前和操作码执行之后。

5. 根据权利要求1所述的基于以太坊虚拟机的代码插桩框架系统，其特征在于，所述的虚拟机在执行过程中的各类数据包括以太坊虚拟机中的栈、内存，交易输入字段、所涉及智能合约的状态变量和交易返回字段、当前交易的交易哈希和所处区块号；

其中，所述的栈、内存指的是以太坊虚拟机中的数据存储结构。

6. 根据权利要求1所述的基于以太坊虚拟机的代码插桩框架系统，其特征在于，所述的数据流分析相关的各类数据指的是动态污点引擎模块自主维护的用于存储和传播污点的影子栈、影子内存、影子交易输入字段、所涉及智能合约的影子状态变量和影子交易返回字段。

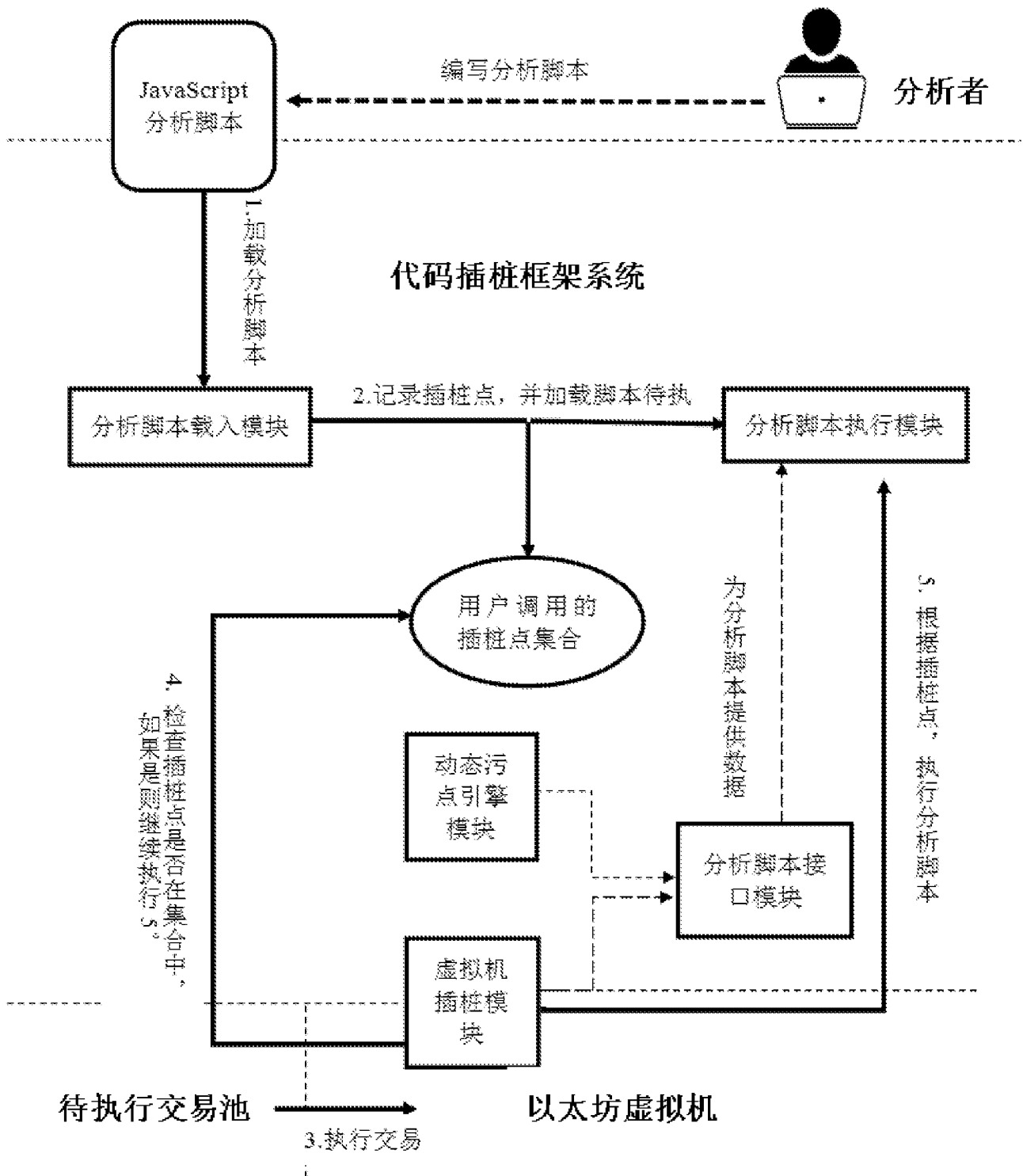


图 1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/095248

A. CLASSIFICATION OF SUBJECT MATTER

G06F 21/54(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06Q; H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC, IEEE: 智能合约, 智能合约, 智能钱包, 木马, 攻击, 交易, 价值, 代码插入, 代码插桩, 程序插桩, 程序插入, 缺陷, 区块链, 漏洞, 病毒, 恶意, 以太坊, 污点, 污染, 外观, 缺陷, 瑕疵, 账户, 外部拥有smart, intelli+, contract +, code, instrumenta+, chink, leak, loophole, blot

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 109800175 A (HOHAI UNIVERSITY) 24 May 2019 (2019-05-24) description paragraphs [0007]-[0035], figures 1-5	1-6
A	CN 110866255 A (BOYA ZHENGLIAN (BEIJING) TECHNOLOGY CO., LTD. et al.) 06 March 2020 (2020-03-06) entire document	1-6
A	CN 111177730 A (HOHAI UNIVERSITY) 19 May 2020 (2020-05-19) entire document	1-6
A	CN 110046089 A (SOUTH CHINA NORMAL UNIVERSITY et al.) 23 July 2019 (2019-07-23) entire document	1-6
A	US 2019034404 A1 (IBM) 31 January 2019 (2019-01-31) entire document	1-6



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

19 February 2021

Date of mailing of the international search report

04 March 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/095248

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109800175	A	24 May 2019	None	
CN	110866255	A	06 March 2020	None	
CN	111177730	A	19 May 2020	None	
CN	110046089	A	23 July 2019	None	
US	2019034404	A1	31 January 2019	US 2019034402 A1	31 January 2019

国际检索报告

国际申请号

PCT/CN2020/095248

A. 主题的分类

G06F 21/54 (2013.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06F; G06Q; H04L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNKI, CNPAT, WPI, EPDOC, IEEE: 智能合约, 智能钱包, 木马, 攻击, 交易, 价值, 代码插入, 代码插桩, 程序插桩, 程序插入, 缺陷, 区块链, 漏洞, 病毒, 恶意, 以太坊, 污点, 污染, 外观, 缺陷, 瑕疵, 账户, 外部拥有 smart, intelli+, contract+, code, instrumenta+, chink, leak, loophole, blot

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 109800175 A (河海大学) 2019年 5月 24日 (2019 - 05 - 24) 说明书第[0007]-[0035段], 附图1-5	1-6
A	CN 110866255 A (博雅整链北京科技有限公司等) 2020年 3月 6日 (2020 - 03 - 06) 全文	1-6
A	CN 111177730 A (河海大学) 2020年 5月 19日 (2020 - 05 - 19) 全文	1-6
A	CN 110046089 A (华南师范大学等) 2019年 7月 23日 (2019 - 07 - 23) 全文	1-6
A	US 2019034404 A1 (IBM) 2019年 1月 31日 (2019 - 01 - 31) 全文	1-6

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2021年 2月 19日

国际检索报告邮寄日期

2021年 3月 4日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

马美红

电话号码 86-(10)-53961516

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/095248

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109800175	A	2019年 5月 24日	无	
CN	110866255	A	2020年 3月 6日	无	
CN	111177730	A	2020年 5月 19日	无	
CN	110046089	A	2019年 7月 23日	无	
US	2019034404	A1	2019年 1月 31日	US 2019034402 A1	2019年 1月 31日