

(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



(10) 国际公布号

WO 2019/242444 A1

(43) 国际公布日
2019 年 12 月 26 日 (26.12.2019) WIPO | PCT

(51) 国际专利分类号:
G06N 99/00 (2019.01)

(21) 国际申请号: PCT/CN2019/087564

(22) 国际申请日: 2019 年 5 月 20 日 (20.05.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810638258.2 2018 年 6 月 20 日 (20.06.2018) CN

(71) 申请人: 深信服科技股份有限公司 (SANGFOR TECHNOLOGIES INC.) [CN/CN]; 中国广东省深圳市南山区学苑大道 1001 号南山智园 A1 栋一层, Guangdong 518055 (CN)。

(72) 发明人: 位凯志 (WEI, Kaizhi); 中国广东省深圳市南山区学苑大道 1001 号南山智园

A1 栋一层, Guangdong 518055 (CN)。章明星 (ZHANG, Mingxing); 中国广东省深圳市南山区学苑大道 1001 号南山智园 A1 栋一层, Guangdong 518055 (CN)。

(74) 代理人: 深圳市深佳知识产权代理事务所 (普通合伙) (SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦 B 座 18C2、18D、18E、18E2, Guangdong 518001 (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, 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,

(54) Title: METHOD AND SYSTEM FOR TRAINING MACHINE LEARNING ENGINE AND RELATED DEVICE

(54) 发明名称: 一种训练机器学习引擎的方法、系统及相关装置

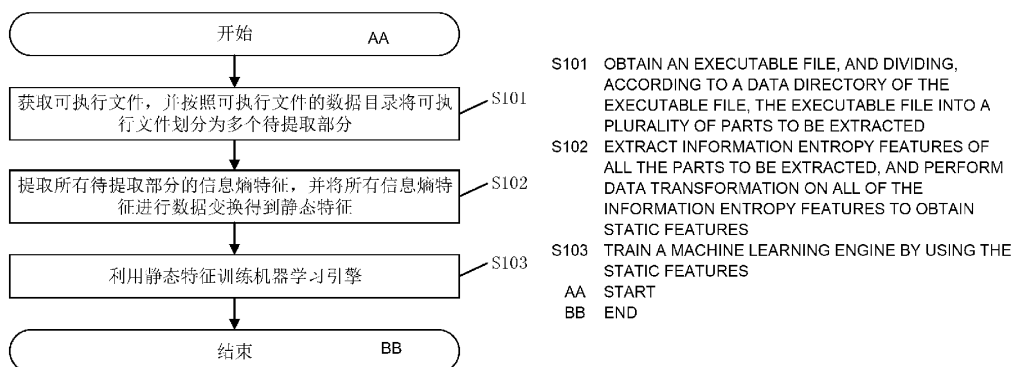


图 1

(57) Abstract: Disclosed are a method and system for training a machine learning engine, a computer readable storage medium and a device for training a machine learning engine, the method comprising: obtaining an executable file, and dividing, according to a data directory of the executable file, the executable file into a plurality of parts to be extracted (S101); extracting information entropy features of all the parts to be extracted, and performing data transformation on all of the information entropy features to obtain static features (S102); and training a machine learning engine by using the static features (S103). The static features of the executable file can be extracted stably so that a detection model generated by the machine learning engine has a strong robust recognition rate.

(57) 摘要: 一种训练机器学习引擎的方法、系统及一种计算机可读存储介质和一种训练机器学习引擎的装置, 该方法包括: 获取可执行文件, 并按照所述可执行文件的数据目录将所述可执行文件划分为多个待提取部分 (S101); 提取所有所述待提取部分的信息熵特征, 并将所有所述信息熵特征进行数据变换得到静态特征 (S102); 利用所述静态特征训练机器学习引擎 (S103)。能够稳定的提取可执行文件的静态特征使得机器学习引擎生成的检测模型具有较强鲁棒性的识别率。

[见续页]

WO 2019/242444 A1

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,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, 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))。