

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
27 February 2020 (27.02.2020)



(10) International Publication Number
WO 2020/040776 A1

- (51) **International Patent Classification:**
G06N 99/00 (2010.01) *G06F 21/55* (2013.01)
- (21) **International Application Number:**
PCT/US2018/047793
- (22) **International Filing Date:**
23 August 2018 (23.08.2018)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant:** **VISA INTERNATIONAL SERVICE ASSOCIATION** [US/US]; 900 Metro Center Boulevard, Foster City, California 94404 (US).
- (72) **Inventors:** **GADDAM, Ajit**; 800 Sea Spray Lane, Apt. 301, Foster City, California 94404 (US). **JERMAKYAN, Ara**; 19012 Plummer Street, Northridge, California 91324 (US). **JOGLEKAR, Pushkar**; 1333 Powell Street, Unit 332, Emeryville, California 94608 (US).

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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

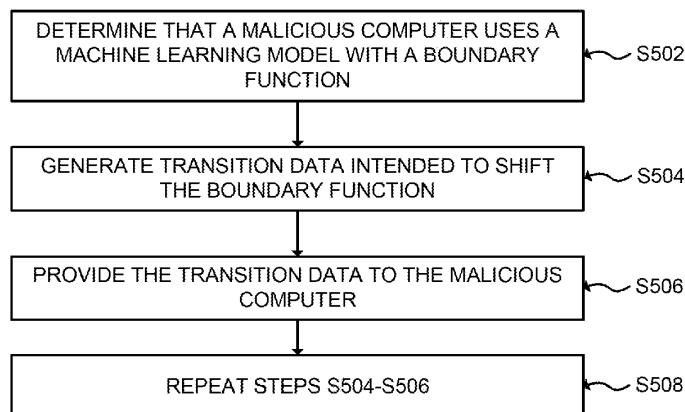
- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, 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).

Published:
— with international search report (Art. 21(3))

- (74) **Agent:** **JEWIK, Patrick** et al.; Kilpatrick Townsend & Stockton LLP, Mailstop: IP Docketing - 22, 1100 Peachtree Street, Suite 2800, Atlanta, Georgia 30309 (US).

- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): 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,

(54) **Title:** PROACTIVE DEFENSE OF UNTRUSTWORTHY MACHINE LEARNING SYSTEM



(57) **Abstract:** Methods and systems for inducing model shift in a malicious computer's machine learning model is disclosed. A data processor can determine that a malicious computer uses a machine learning model with a boundary function to determine outcomes. The data processor can then generate transition data intended to shift the boundary function and then provide the transition data to the malicious computer. The data processor can repeat generating and providing the transition data, thereby causing the boundary function to shift over time.

WO 2020/040776 A1