



(51) **International Patent Classification:**
G06F 21/55 (2013.01) *G06F 18/2115* (2023.01)
G06N 20/10 (2019.01)

(21) **International Application Number:**
PCT/EP2024/075366

(22) **International Filing Date:**
11 September 2024 (11.09.2024)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
63/582,278 13 September 2023 (13.09.2023) US
18/401,364 30 December 2023 (30.12.2023) US

(71) **Applicant: BITDEFENDER IPR MANAGEMENT LTD [CY/CY];** 59-61 Acropolis ave. 3rd floor, office 302, Nicosia, 2012 (CY).

(72) **Inventors: SMEU, Stefan;** Str. Corbului nr. 17D, Județul Bacău, 600401 Bacău (RO). **BURCEANU, Elena;** Str. Constantin Radulescu-Motru nr. 4, bl. 1, sc. 2, et. 6, ap. 58, Sectorul 4, 040361 București (RO). **HALLER, Emanuela;** Str. Preciziei nr. 6M, Bl. 9B, Ap. 23, Sectorul 6, 110336 București (RO).

(74) **Agent: CABINET DOINA TULUCA;** Bd. Lacul Tei Nr. 56, bl. 19. sc. B, ap. 52, sector 2, 020392 București (RO).

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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).

(54) **Title:** FEATURE SELECTION USING FEATURE DISTRIBUTION IN DIFFERENT SUBSETS FOR MACHINE LEARNING IN COMPUTER SECURITY APPLICATIONS

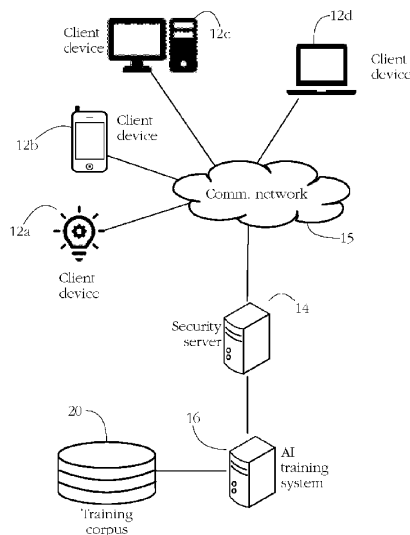


FIG. 1

(57) **Abstract:** A diverse collection of data samples harvested for computer security applications is divided into a plurality of training corpora according to criteria such as an identity of the data source, among others. An initial set of features for characterizing the collected data samples is reduced to an optimal subset. For each candidate feature, a frequency distribution of feature values is determined over each of the training corpora. A feature selection procedure favors features whose frequency distribution is relatively similar across multiple corpora. A detector module is then trained to detect computer security threats according to the reduced set of features.

Published:

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