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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 September 2007 (07.09.2007)

PCT

(10) International Publication Number
WO 2007/100916 A3

(51) International Patent Classification:
G06F 12/14 (2006.01)

(21) International Application Number:
PCT/US2007/005408

(22) International Filing Date:
28 February 2007 (28.02.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/778,008	28 February 2006 (28.02.2006)	US
60/790,626	10 April 2006 (10.04.2006)	US

(71) Applicant *(for all designated States except US)*: **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; 412 Low Memorial Library, 535 West 116th Street, New York, NY 10027 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **STOLFO, Salvatore, J.** [US/US]; 80 Kenilworth Road, Ridgewood, NJ 07450 (US). **WANG, Ke** [CN/US]; 435 W. 119th Street, Apt. 2E, New York, NY 10027 (US). **PAREKH, Janak** [US/US]; 110 Bayview Road, Manhasset, NY 11030 (US).

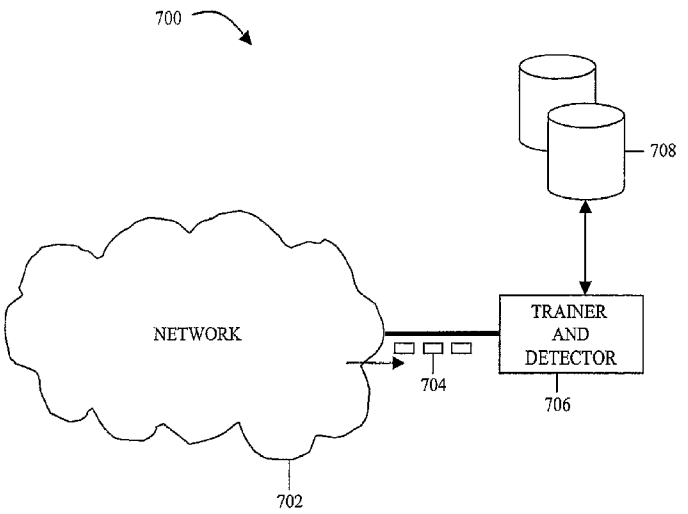
(74) Agents: **BYRNE, Matthew, T.** et al.; Wilmer Cutler Pickering Hale and Dorr LLP, 399 Park Avenue, New York, NY 10022 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: SYSTEMS, METHODS, AND MEDIA FOR OUTPUTTING A DATASET BASED UPON ANOMALY DETECTION



(57) Abstract: Systems, methods, and media for outputting a dataset based upon anomaly detection are provided. In some embodiments, methods for outputting a dataset based upon anomaly detection: receive a training dataset having a plurality of n-grams, which plurality includes a first plurality of distinct training n-grams each being a first size; compute a first plurality of appearance frequencies, each for a corresponding one of the first plurality of distinct training n-grams; receive an input dataset including first input n-grams each being the first size; define a first window in the input dataset; identify as being first matching n-grams, the first input n-grams in the first window that correspond to the first plurality of distinct training n-grams; compute a first anomaly detection score for the input dataset using the first matching n-grams and the first plurality of appearance frequencies; and output the input dataset based on the first anomaly detection score.

WO 2007/100916 A3

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
24 April 2008

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 07/05408

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 12/14 (2007.01)
USPC - 726/22
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)
726/22

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 714/100; 714/709; 717/124

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST - terms: output dataset, anomaly, anomaly detection, training dataset, n-gram, secret basis, secret, random, pseudo-random, pseudo random, count, hash function, processor, match, matching, network data, probability, classify, classification, malicious, malicious code, consistency, score, predetermine, etc ...

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2003/0182310 A1 (CHARNOCK, et al.) 25 September 2003 (25.09.2003), abstract and para [0066], [0068]-[0072], [0099]-[0101], [0139], [0150], [0191], [0204], [0228], [0233], [0246], [0248], [0252]-[0253], [0297], [0304]-[0305], [0365], [0445], [0449], [0453]-[0455], [0457], [0481], [0495], [0500]-[0503], [0511], [0513]-[0514], [0533], [0550], [0661], [0707], [0723], [0730], [0733], [0735], [0741], [0846], [0857], [0862], [0868], [0873], [0887], [0893].	1-4, 6, 8, 11-15, 18-35, 37, 39, 42-46, 49-66, 68, 70, 73-77 and 80-93 --- 5, 7, 9-10, 16-17, 36, 38, 40-41, 47-48, 67, 69, 71-72 and 78-79
Y	US 2004/0111410 A1 (BURGOON, et al.) 10 June 2004 (10.06.2004), abstract and para [0161]-[0165].	5, 36 and 67
Y	US 2006/0015630 A1 (STOLFO, et al.) 19 January 2006 (19.01.2006), abstract and para [0091]-[0093].	7, 16-17, 38, 47-48, 69 and 78-79
Y	US 2005/0249214 A1 (PENG) 10 November 2005 (10.11.2005), abstract and para [0071].	9-10, 40-41 and 71-72
A	US 2003/0014662 A1 (GUPTA, et al.) 16 January 2003 (16.01.2003), entire document, especially abstract and para [0042]-[0048], [0050]-[0058], [0069]-[0076], [0081]-[0084], [0164]-[0168].	1-93

☐ Further documents are listed in the continuation of Box C. ☐

* 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	Date of mailing of the international search report
22 August 2007 (22.08.2007)	06 MAR 2008
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)