



- (51) International Patent Classification:
G06F 17/30 (2006.01) *G06F 15/16* (2006.01)
- (21) International Application Number:
PCT/US2011/066375
- (22) International Filing Date:
21 December 2011 (21.12.2011)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
12/981,393 29 December 2010 (29.12.2010) US
12/981,397 29 December 2010 (29.12.2010) US
- (71) Applicant (for all designated States except US):
AMAZON TECHNOLOGIES, INC. [US/US]; P.O. Box 8102, Reno, Nevada 89507 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **SORENSEN III, James, Christopher** [US/US]; 1200 12th Avenue South, Suite 1200, Seattle, Washington 98144-2734 (US).
- (74) Agent: **MEYERTONS, HOOD, KIVLIN, KOWERT & GOETZEL, P.C.**; KOWERT, Robert C., P.O. Box 398, Austin, Texas 78767-0398 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, 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, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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,

[Continued on next page]

(54) Title: RECEIVER-SIDE DATA DEDUPLICATION IN DATA SYSTEMS

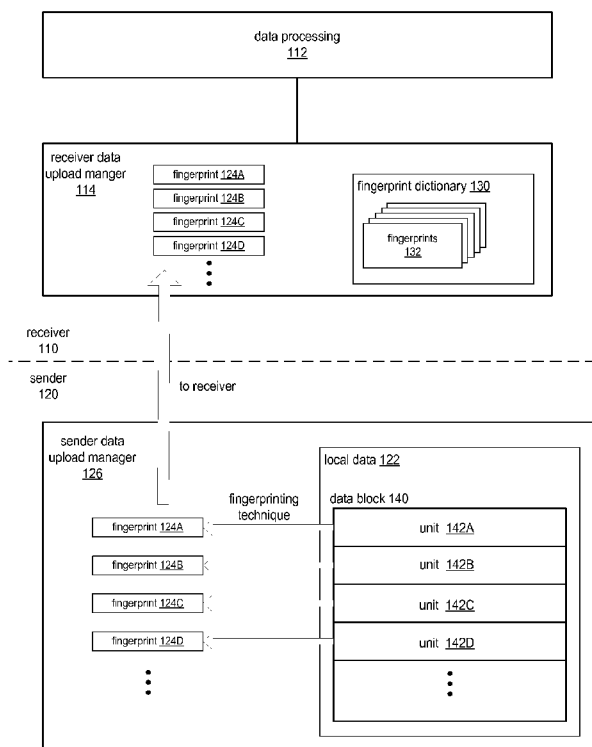


Figure 2A

(57) Abstract: Methods and apparatus for receiving uploaded data from a sender at a receiver. A data deduplication technique is described that may reduce the bandwidth used in uploading data from the sender to the receiver. In the technique, the receiver, rather than the sender, maintains a fingerprint dictionary for previously uploaded data. When a sender has additional data to be uploaded, the sender extracts fingerprints for units of the data and sends the fingerprints to the receiver. The receiver checks its fingerprint dictionary to determine the data units to be uploaded and notifies the sender of the identified units, which then sends the identified units of data to the receiver. The technique may, for example, be applied in virtualized data store systems to reduce bandwidth usage in uploading data.



SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished
upon receipt of that report (Rule 48.2(g))*