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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 June 2016 (16.06.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/091278 A1

(51) International Patent Classification:
H04L 29/08 (2006.01) *H04W 4/00* (2009.01)

(21) International Application Number:
PCT/EP2014/076899

(22) International Filing Date:
8 December 2014 (08.12.2014)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **NEC EUROPE LTD.** [DE/DE]; Kurfürsten-Anlage 36, 69115 Heidelberg (DE).

(72) Inventors: **PAPAGEORGIOU, Apostolos**; Pfaffengasse 9, 69117 Heidelberg (DE). **CHENG, Bin**; Finkenweg 25, 69214 Eppelheim (DE). **KOVACS, Ernoe**; Kleinhohenheimerstraße 20, 70619 Stuttgart (DE).

(74) Agent: **ULLRICH & NAUMANN**; Schneidmühlstraße 21, 69115 Heidelberg (DE).

(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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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))*

(54) Title: METHOD AND SYSTEM FOR FILTERING DATA SERIES

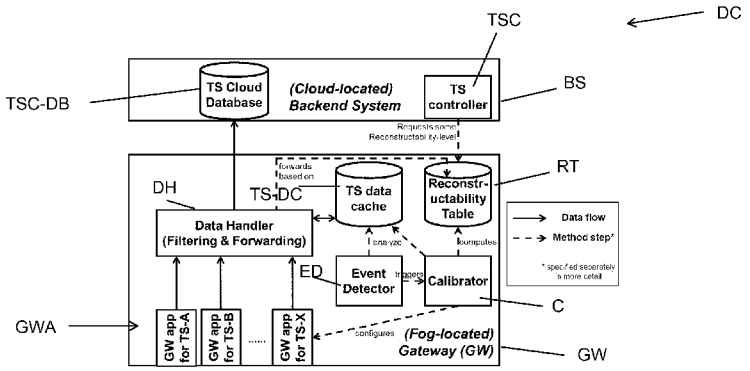


Fig. 2

(57) Abstract: The present invention relates to a method for filtering data series, preferably time series of data, prior to further processing, wherein the data series are collected by collecting entities and provided to one or more filtering entities from one or more data delivering devices, and wherein the filtered data series are forwarded to further processing entities, wherein the filtering by said filtering entities is performed by the steps of a) Collecting a data series, b) Reducing the information of the data series based on and by at least one data reduction procedures, c) Reconstruction of the original information for each reduced information of a data series, d) Calculating the level of reconstruction for the information based on a comparison between the reconstructed information and the original information collected in step a) for at least one of the reduction procedures, and e) Determining the reduced or non-reduced information of the data series to be forwarded based on a comparison between a desired level and the calculated level of reconstruction.

WO 2016/091278 A1

METHOD AND SYSTEM FOR FILTERING DATA SERIES

The present invention relates to a method for filtering data series, preferably time series of data, prior to further processing, wherein the data series are collected by collecting entities and provided to one or more filtering entities from one or more data delivering devices, and wherein the filtered data series are forwarded to further processing entities.

The present invention further relates to a system for filtering data series, preferably time series of data, prior to further processing, comprising one or more data delivering devices adapted to provide data series, one or more collecting entities adapted to collect said data series and to provide them to one or more filtering entities and wherein said one or more filtering entities are adapted to forward the filtered data series to further processing entities.

Although applicable to any kind of systems, the present invention will be described with regard to data series in connection with the internet-of-things or machine-to-machine systems.

Although applicable in general to any type of data series, the present invention will be described with regard to time series of data.

In internet-of-things or machine-to-machine systems devices conventionally send or actuate or any other automated data-generating task constantly provide information about any object, also called “thing”, mostly in form of so-called time series. Time series usually refer to data that are generated and/or collected at successive times in regular or irregular intervals and comprise a key-value pair. For example the value is a simple data type, for instance numeric, alphanumeric or binary data and a corresponding timestamp. For example time series stemming from internet-of-thing devices are one of the enablers of the so-called big data.

Time series collected by internet-of-thing devices are often forwarded and stored via deployments based on a system illustrated in Fig. 1. For instance the data provided by M2M devices D goes either via a cellular network or via proxies like

edge routers or gateway devices GW and through a backbone network NC to a backend system BS storing, processing and offering related information for example with a common application programming interface API to applications of various domains.

Conventionally the data is forwarded and stored in a data center DC, for example a cloud. However, this causes a plurality of problems: For instance one of the problems is the bandwidth consumption and/or latency between the data delivering devices D or the gateways GW and the network core NC or the data center DC respectively. The further problems are the storage costs and the database performance of a data center DC. Another problem is the energy consumption on various tiers and further the system resilience because of potential concurrent database transactions. With the increasing use of the internet-of-things these problems will become even bigger in the future.

To address these problems the non-patent literature of Tak-chung Fu, “A review on time series data mining”, Engineering Applications of Artificial Intelligence, Volume 24, Issue 1, February 2011, Pages 164-181, ISSN 0952-1976 and of W. Lang, M. Morse, and J. M. Patel, "Dictionary-Based Compression for Long Time-Series Similarity," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 11, pp. 1609-1622, November, 2010 apply conventional reduction procedures like sampling, compression and/or selected forwarding, for example rule based and/or application-specific, or are customized for conventional internet-of-thing architectures such as for example shown in Fig. 1 or for important internet-of-things application domains such as transportation, industrial automation, safety, etc..

In the further non-patent literature of J. Zhang, K. Yang, L. Xiang, Y. Luo, B. Xiong, and Q. Tang, “A Self-Adaptive Regression-Based Multivariate Data Compression Scheme with Error Bound in Wireless Sensor Networks”, International Journal of Distributed Sensor Networks, Vol. 2013, Article ID 913497 a method is shown for deciding automatically to transmit either raw or regression coefficients and in the latter case to select the number of data involved in the regressions.

However, these conventional methods act upon already collected data sets. Further they are often avoided because of the information loss that selected forwarding or data filtering inherently applies.

In Fig. 1 this effect is illustrated with different data reduction procedures F1, F2, F3 applied on the original collected data series O-TS. For example when on the original time series with data collected in period T2 the data reduction method F1 is applied then the filtered data is completely lost, indicated by a non-present bar. The same is true for data reduction method F2. With data reduction method F3 a smaller amount of data is available afterwards. For the time series collected in period T4 the same is true for the data filtered with the data reduction procedure F1 and with the sampling according to data reduction mechanism F2 whereas when compression F3 is applied on the data collected in period T4 the compression has no effect, indicated by a non-changed bar in Fig. 1.

These information “losses” are very difficult to determine, especially when designing a data-agnostic system, i.e. a system that cannot filter based on the semantics of the data or based on application-specific needs. One reason for example is that it is unknown who will use the data and in which way.

It is therefore an objective of the present invention to provide a method and a system for filtering data series with enhanced efficiency in terms of storage, bandwidth and averaged data quality preferably in an internet-of-things or machine-to-machine system.

It is a further objective of the present invention to provide a method and a system for filtering data series which can maintain the desired level for the reconstructability of the original data from the subset that has been forwarded and/or stored in a data center.

It is an even further objective of the present invention to provide a method and a system for filtering data series enhancing a control of the degree of information “loss” due to filtering independently of which filtering method is being applied.

The aforementioned objectives are accomplished by a method of claim 1 and a system of claim 12.

In claim 1 a method for filtering data series, preferably time series of data, prior to further processing, wherein the data series are collected by collecting entities and provided to one or more filtering entities from one or more data delivering devices, and wherein the filtered data series are forwarded to further processing entities is defined.

According to claim 1 the method is characterized in that the filtering by said filtering entities is performed by the steps of

a) Collecting a data series,
b) Reducing the information of the data series based on and by at least one data reduction procedures,
c) Reconstruction of the original information for each reduced information of a data series,
d) Calculating the level of reconstruction for the information based on a comparison between the reconstructed information and the original information collected in step a) for at least one of the reduction procedures, and
e) Determining the reduced or non-reduced information of the data series to be forwarded based on a comparison between a desired level and the calculated level of reconstruction.

In claim 12 a system for filtering data series, preferably time series of data, prior to further processing, comprising one or more data delivering devices adapted to provide data series, one or more collecting entities adapted to collect said data series and to provide them to one or more filtering entities and wherein said one or more filtering entities are adapted to forward the filtered data series to further processing entities is defined.

According to claim 12 the system is characterized in that said one or more filtering entities are adapted to perform the steps of

a) Collecting a data series,

b) Reducing the information of the data series based on and by at least one data reduction procedures,
c) Reconstruction of the original information for each reduced information of a data series,
d) Calculating the level of reconstruction for the information based on a comparison between the reconstructed information and the original information collected in step a) for at least one of the reduction procedures, and
e) Determining the reduced or non-reduced information of the data series to be forwarded based on a comparison between a desired level and the calculated level of reconstruction.

The term “reconstructability” is to be understood preferably in the claims, in particular in the description as a degree to which an original data set can be reproduced from a reduced instance of the data set, preferably the original data set with missing points or values, the function that can be used to retrieve points of the data set, or a different data set that has been generated from a transformation of the original data set. It is preferably expressed as a percentage, based on system-specific matrix, etc..

The term “gateway” is to be understood preferably in the claims, in particular in the description in its broadest sense, in particular to be understood as entity at a network edge.

The term “entity” is to be understood preferably in the claims, in particular in the description in its broadest sense, in particular an entity like a filtering entity can preferably also act as further processing entity and/or act as entity of another type, etc.

According to the invention it has been recognized that it can be easily determined which data to forward, which to filter and which to cache based on a reconstructability of the data series, preferably of data points of time series.

According to the invention it has been further recognized that the efficiency in terms of storage, bandwidth and average data quality is enhanced, while simultaneously maintaining a predefined level for the reconstructability of the original data from the subset that has been stored, for example in a data center.

According to the invention it has been recognized that filtering or compression techniques for time series can preferably applied before actually collecting and are replied upon preferably frontend samples.

According to the invention it has been even further recognized that by using reconstructability levels settings of the data reduction procedures are translated into degrees of information loss.

According to the invention it has been even further recognized that knowledge about the reconstructability of the data series is related with decisions about settings of used data reduction procedures.

According to the invention it has been even further recognized that data-agnostic data filtering with controlled degree of information loss is enabled.

In other words the present invention provides a translation of settings of data reduction procedures into degrees of expected information losses. Further the present invention relates the knowledge about reconstructability of data with decisions about settings of the used data reduction procedures. Further the present invention enables data agnostic data filtering with a controlled degree of information loss.

Further features, advantages and preferred embodiments are described in the following subclaims.

According to a preferred embodiment at least steps a)-d) are performed in irregular and/or regular time intervals, upon prespecified changes and/or appearance of prespecified values in data series. This enables in a simple and efficient way to trigger filtering and the analysis of incoming data series: When and

how frequently the data is being (re)examined is determined. A simple timer may trigger the analysis in regular or irregular intervals. An event detector may trigger the analysis upon detection that certain prespecified values of the data series are changing and/or are exceeded a certain prespecified threshold. Another possibility is that the event detector may trigger the analysis upon appearance of certain prespecified values in the information of the data series that indicate a change of behavior. Of course any other procedure may alternatively or additionally be used.

According to a further preferred embodiment when collecting the data series the highest possible polling rate and/or the highest possible resolution is used. This enables to provide most actual and/or most precise data when collecting the data series, for example based on the available bandwidth of the communication between the data delivering devices and the filtering entities. Further precision of the reduction procedures is enhanced since the largest possible amount of data for later analysis can be used.

According to a further preferred embodiment reconstructability information is generated specifying for each data series and for each reduction procedure and for corresponding input values for said reduction procedures a value for the level of reconstruction. This enhances the flexibility to a great extent which reduced data shall be forwarded for further processing to the further processing entities.

According to a further preferred embodiment the reconstructability information are updated when steps a)-d) are performed. This enables providing most actual reconstructability information for deciding which data to be forwarded in what way.

According to a further preferred embodiment a reduction procedure is provided in form of a procedure reducing dimensionality and/or size of the data series and/or a generation of a function representing the data series. Dimensionality reduction is for example provided in sampling of each, every second, every fourth or no data point of a data series. Function-based representation of a reduction procedure, for example forwards only a function which represents the data “as good as possible”, for example only every second data point is used and a spline function is generated through every second data point and the function of said spline together

with the corresponding data interval is forwarded for further processes providing efficient reduction procedures. Of course any other data reduction procedure can be used additionally or alternatively. Also applying of different reduction procedures sequentially is possible.

According to a further preferred embodiment the comparison according to step d) is performed on a similarity metric, preferably using Euclidian distance. This enables in a fast and efficient way to provide a comparison between the reconstructed information and the original information.

According to a further preferred embodiment the collecting entities are configured based on the operational status of said filtering entities. This enhances the flexibility while providing an optimum of communication between the filtering entities and the collecting entities.

According to a further preferred embodiment when the operational status of the filtering entities is dedicated for energy saving then the collecting entities are reconfigured such that only reduced information satisfying the desired level of reconstruction is collected. “Reduced” means here as much as needed to satisfy the reconstructability degree that has been requested. This reduces the collecting entity traffic and saves energy of the collecting entity.

According to a further preferred embodiment when the operational status of the filtering entities is dedicated for network resource saving then the collecting entities are reconfigured such that only reduced information satisfying the desired level of reconstruction are forwarded and preferably the collected information is cached in the filtering entity and/or in the collecting entity. This releases the network and storage demand equally to the reconfiguration of the collecting entities for energy saving and keeps more collected data in the cache which might be retrieved later. Therefore, the flexibility is further enhanced since data in the cache can be provided at any time if needed.

According to a further preferred embodiment when the operational status of the filtering entities is dedicated for network resource saving, the collected information

are forwarded upon demand of the further processing entities in regular time intervals and/or never. “Upon demand” means that data can be eventually retrieved upon request. This preferably means that it may take time until the data is delivered for example to optimize bandwidth usage or because intermediate nodes are unreliable such that manual fetching of the data to the backend system is preferred. Another option is that big amounts of data will be sent at all to the backend system if not explicitly requested for instance. “In regular time interval” means that the data cached is copied, or i.e. transmitted to the backend system BS regularly with time intervals that are preferably much bigger than the data capture intervals. If they are forwarded never then the data series cached can only be used locally and might be dropped at any time.

There are several ways how to design and further develop the teaching of the present invention in an advantageous way. To this end it is to be referred to the patent claims subordinate to patent claim 1 on the one hand and to the following explanation of preferred embodiments of the invention by way of example, illustrated by the figure on the other hand. In connection with the explanation of the preferred embodiments of the invention by the aid of the figure, generally preferred embodiments and further developments of the teaching will be explained. In the drawings

Fig. 1 shows a conventional internet-of-thing deployment;

Fig. 2 shows a part of a system according to a first embodiment of the present invention;

Fig. 3 shows a part of a method according to a second embodiment of the present invention;

Fig. 4 shows a part of a system according to a third embodiment of the present invention;

Fig. 5 shows a part of a method according to a fourth embodiment of the present invention; and

Fig. 6 shows a part of a system according to a fifth embodiment of the present invention.

Fig. 1 shows a conventional internet-of-thing deployment.

In Fig. 1 time series for filtering in a conventional internet-of-thing deployment is shown. A number of embedded systems and sensors D is connected to a multi-service edge like gateways, edge routers or the like GW. These gateways GW collected data from the devices D which is illustrated by the table indicating as bars the data within the time periods T1, T2, T3 in original time series O-TS. The gateways GW are connected via a network core NC to a backend system BS comprising a data center DC. The gateways GW provide filtered data of the original time series O-TS which is depicted on the upper right corner of Fig. 1. The data within the periods T1, T2, T3 are then transmitted to the data center DC reduced with some filtering F1 procedure or some sampling rate F2 or some compression procedure F3. The bars in the table of the time series indicate the level or the amount of data after the data reduction procedure F1, F2, F3 has been applied. For example for the data series in period T1 and the filtering F1 the data series in period T1 of the original data series O-TS corresponds to the filtered one whereas for said original data series in period T1 on which the compression mechanism F3 has been applied smaller data was transmitted to the data center DC, as depicted in Fig. 1 with a smaller bar.

Fig. 2 shows a part of a system according to a first embodiment of the present invention.

In Fig. 2 a system for the enablement of reconstructability-aware time series handling is shown.

In Fig. 2 a gateway GW is shown comprising gateway applications GWA for different time series TS-A, TS-B, ..., TS-X. These gateways applications GWA are connected to a data handler DH for filtering and forwarding of data series provided

to the data handler DH by the gateway applications GWA. The data handler DH is further connected to a backend system BS to forward filtered data comprising a time series cloud database TSC-DB. Further the backend system BS comprises a time series controller TSC which requests some reconstructability level from a reconstructability table RT which is again located or stored in the gateway GW. Further the gateway GW comprises a time series data cache TS-DC, an event detector ED and a calibrator C. The event detector ED triggers the calibrator C to analyze the data and to update the reconstructability table RT. The data handler DH exchanges data with the time series data cache TS-DC. When the event detector ED triggers the calibrator C the calibrator C preferably configures the gateway applications GWA.

The above mentioned entities in the gateway GW and the backend system BS are in the following described in more detail:

- Gateway application GWA: This entity may be preferably realized in form of a software module that communicates directly with the data source, i.e. with an M2M/IoT device D. The pattern of its interaction with the data source D can be configured by another module, but the Gateway Application GWA itself has no knowledge about the actual data reduction procedures.

- Data Handler DH: This entity offers for example an API to the gateway applications GWA, through which the latter can report the data they receive from the data sources D. The data handler DH is then responsible for writing them back to the Cloud Database TSC-DB, to the local Cache TS-DC or both.

- Time Series Data Cache TS-DC: This entity may be provided as a cache memory of the gateway GW which stores time series, if possible in exactly the same way they are stored in the Cloud database TSC-DB. The purpose of caching may be threefold: a) use cache data sets to measure the statistical characteristics of the data that are related to their reconstructability, and/or b) keep a copy of data that have not been stored in the Cloud, and/or c) keep (recent) data close to the data sources D for

cases where fast or "offline" response/actuation is required. From the above, it is preferably "a)" that is directly related to the description of method steps that will follow.

- Event Detector ED: This entity looks into the incoming/cached time series TS and determines if and when a statistical analysis of the data is required in order to adjust the reconstructability table RT, i.e. the knowledge base based on which the forwarding/filtering of the Data Handler DH is done. The Event Detector ED can operate in parallel with the Data Handler DH and independently of the data forwarding process. Its operation does not affect the flow of the actual data until a new reconstructability table RT has been calculated.

- Calibrator C: This entity may be triggered by the Event Detector ED in order to do the analysis of the cached data and update the reconstructability table RT, as well as the configuration of the Gateway Applications GWA.

- Reconstructability table RT: This entity is a knowledge base comprising all the information that the Data Handler DH needs in order to decide how to filter and forward data, e.g. maximum acceptable degree of information loss, mapping of data reduction procedures, settings to degrees of information loss, and more.

- Time Series Controller TSC: This entity - among others - informs the gateway GW about the maximum acceptable degree of information loss.

- Time Series Cloud Database TSC-DB: This is the Cloud Database where the time series data is stored.

Fig. 3 shows a part of a method according to a second embodiment of the present invention.

In Fig. 3 a high level flow of the reconstructability-aware time series forwarding and filtering procedure is shown.

In a first phase P1 the time series is analyzed with steps S1.1-S1.3 and a second phase P2 filters and forwards the data with steps S2.1 and S2.2, wherein both phases P1, P2 may be at least partially performed in parallel. The steps are now described in more detail:

Step S1.1

The Event Detector ED triggers an analysis of incoming time series, for example upon fulfillment of a custom condition. For triggering this, the Event Detector ED has a mechanism or procedure which determines when and how frequently the data is being re-examined in order to update the reconstructability table RT. This mechanism/procedure can be for example:

- A simple timer, which triggers the analysis in regular or irregular intervals.
- An event detector ED, which triggers the analysis upon the detection of the fact that the changes of certain (pre-specified) values of the time series are exceeding a certain (pre-specified) threshold.
- An event detector ED, which triggers the analysis upon the appearance of certain (pre-specified) values in the Time Series data which commonly indicate a change of behavior and/or
- Any mechanism similar to the above.

Step S1.2

When the Calibrator C is triggered by the Event Detector ED (Step S1.1), then the following sub-steps are preferably performed:

- The GW applications GWA that communicate with the devices/sensors D are reconfigured by the Calibrator C so that for the monitored time series they use the:
  - Highest possible polling rates and/or
  - Highest possible quality/resolution, if, e.g. media content is involved.
- The GW applications GWA collect data with the above configuration for a period T1 and write it into the cache TS-DC
  - The length of period T1 is pre-set or

- The length of period T1 is domain- and time series-dependent and is expected to vary, e.g. between a few seconds and various hours or even days.

Step S1.3

Upon expiration of the time period T1, the Calibrator C uses the data collected during T1 to compute the reconstructability table RT.

Step S2.1

Once the reconstructability table RT has been computed, two options may be performed:

- If the gateway GW has an "energy-saving-mode", i.e. an operational state with reduced energy consumption, being activated then the Calibrator C re-configures the Gateway applications GWA to retrieve only the "reduced" data from the data sources. „Reduced" means as much as needed to satisfy the reconstructability degree that has been requested, e.g., in Fig. 5 when performing Step 1.3 half of the data points (i.e., use "1:2-dimensionality-reduction") are sent if a reconstructability degree of 90% had been indicated as sufficient by the TS controller TSC. The indication of a "required reconstructability degree" by the TS controller TSC may be performed by any suitable means. It is assumed that the "reconstructability requirement" has been given by the TS controller TSC to the gateway GW asynchronously at some point before Phase P2.) This energy saving mode reduces device traffic and saves device energy, but captures less data.

- If the the gateway GW has a "network-relieving-mode", i.e. an operational state with reduced data transmission, being activated then the gateway GW applications GWA will continue retrieving data the same way as during the period T1, but the Data Handler DH will send only the "reduced" data to the backend BS and keep the rest only in the TS data cache TS-DC. This option can be more heavyweight for the devices, but relieves the network and storage demand equally to the first option and keeps more data in the

cache TS-DC, which might be retrieved later. Given that this mode keeps all data points in the cache, the behavior of the cache must also be specified.

Therefore, the "network-relieving-mode" has preferably three sub-modes:

- "on-demand": The data points of the TS data cache TS-DC can be eventually retrieved upon request of the TS controller TSC. This means that it might take time until the data is delivered to the Cloud BS, e.g. to optimize bandwidth usage, or because the intermediate nodes are unreliable, so manual fetching of the data to the Cloud BS is preferred or that big amounts of data will not be sent at all to the Cloud BS, i.e., if not explicitly requested.
- "regular-update": This functions in essence like the "on-demand" option, but the TS data cache TS-DC is copied to the Cloud BS regularly, with time intervals that are typically much bigger than the data capture intervals, e.g., once a day. Thus, this sub-mode is preferably applied when cloud storage efficiency is not an issue because all data is stored there eventually and no guarantees about data availability and resilience are required, because the TS data cache TS-DC is less reliable than the Data Center for instance.
- "never": TS data cache TS-DC can be used only locally and might be dropped at any time.

In this case the gateway GW can be preferably operating either in the "energy-saving-mode" or in the "network-relieving-mode".

Step S2.2

Step S2.1 is preferably never interrupted, but it is dependent on the reconstructability table RT and on further system configuration settings, which can be modified when a new iteration of the entire Phase P1 takes place, triggered in Step S2.2.

Fig. 4 shows a part of a system according to a third embodiment of the present invention.

In Fig. 4 a visualization of the reconstructability table RT is shown. The following is assumed:

- The existence of X incoming Time Series, e.g., collected/reported from sensors, cameras, smartphones etc. The set of Time Series is defined as $TS := (TS_1, TS_2, \ldots, TS_X)$
- The existence of Y data reduction procedures, e.g., “sampling”, “selective filtering”, “compression”. The set of data reduction procedures is defined as $RM := (RM_1, RM_2, \ldots, RM_Y)$.
- The existence of n_K applicable values for RM_K. The set of applicable values for RM_K is defined as $V_K := (val_{0,K}, val_{1,K}, \ldots, val_{(nK-2),K}, val_{100,K})$, whereby $val_{0,K}$ is the value for which no data at all is forwarded In this case the reconstructability degree is defined to be equal to 0%, although a “random reconstruction” could give values that have a similarity >0 with the original data set and $val_{100,K}$ is the value for which the reconstructability degree is equal to 100%, while the rest of the values lead to reconstructability degrees between 0% and 100% (including 0 and 100).

Then, the reconstructability table RT is computed as follows: For each triple (t, r, v) where $t \in TS$, $r \in RM$, and $v \in V_1 \cup V_2 \cup \ldots \cup V_Y$, i.e., for each combination of a time series with a data reduction procedure and a value of this data reduction procedure the reconstructability degree ρ of the triple is measured. The computation of ρ can be based, for example, on the Euclidean distance between the vector of the original data and the vector of the reconstructed data. Similarly, the reconstruction might be performed with linear interpolation or any similar method. ρ is calculated as the degree to which the data of time series t that was collected during period T1 can be reconstructed after it has been reduced with method r using the value v.

Fig. 5 shows a part of a method according to a fourth embodiment of the present invention.

In Fig. 5 a data reduction and reconstruction with various applicable values of two reduction methods, i.e. dimensional reduction reduction and function-based

representation is shown. In Fig. 5 an original time series O-TS also named TS_1 with a plurality of values V is incoming during the period T1. For example the values V can be smart meter values measured over time.

Further two reduction procedures will be used:

- RM_1: Dimensionality reduction (i.e., sampling of each, every second, every fourth, or no data point of TS_1). Thus, V_1 := (0:∞, 1:4, 1:2, 1:1).
- RM_2: Function-based representation, i.e., forwarding to the Cloud only a function, which represents the data "as good as possible". Here it is assumed that two such functions come into question for TS_1, namely f(x) and g(x). Thus, V_2 := (0:∞, f(x), g(x), TS_1-func) whereby 0:∞ forwards nothing to the Cloud, while "TS_1-function" is defined as the function that gives exactly all the data points of TS_1. The latter is, of course, not always possible in reality, and even if it is, it might require a representation that is similar to or bigger than the full representation of the actual data.

Thus, the middle row of graphs of Fig. 5 shows the reduced (circular) and the reconstructable (triangle) data points of TS_1 when RM_1 is applied with its four different applicable values, while the lower row of graphs of Fig. 5 shows the data that is forwarded when RM_2 is applied with its four different applicable values.

Now, the Calibrator C:

- calculates a reconstructed version of TS1 (partly shown in Fig. 5) for each of the eight cases of reduced data (middle and lower row of Fig. 5),
- compares each reconstructed version with the original data set (upper graph of Fig. 5) based on a similarity metric, e.g., Euclidean distance, thus computing the so-called reconstructability degree or level and
- writes the computed reconstructability degree into the reconstructability table RT.

In this example, it is assumed that the computed reconstructability degrees for 1:2-dimensionality-reduction and 1:4-dimensionality-reduction were 95% and 55%,

respectively, while the reconstructability degrees for f(x) and g(x) were 80% and 70%, respectively:

- $\rho(val_{0,1}) = 0$
- $\rho(val_{1,1}) = 55$
- $\rho(val_{2,1}) = 95$
- $\rho(val_{100,1}) = 100$
- $\rho(val_{0,2}) = 0$
- $\rho(val_{1,2}) = 80$
- $\rho(val_{2,2}) = 70$
- $\rho(val_{100,2}) = 100$.

Fig. 6 shows a part of a system according to a fifth embodiment of the present invention.

In Fig. 6 an example instance of a reconstructability table RT is shown based on the values of Fig. 5. With a time series TS the corresponding reduction mechanism RM and the reconstructability level RCL corresponding to reconstruction values RCV of the corresponding reduction mechanism RM.

In summary the present invention enables determination which data to forward, which to filter and which to cache based on the reconstructability of time series data points. Further the present invention enables using time series compression procedures or techniques before the time series are actually collected upon a frontend samples. The present invention further enables to apply a phase-change procedure based on an analysis of data streams comprising a calibration phase/operation phase and to trigger by the main specific events captures with the local data analytics.

The present invention preferably provides a method for filtering and forwarding of time series data in an internet-of-things environment based on data-reconstructability metrics comprising the steps of:

1) Reception of a required reconstructability degree from the backend to the network edge devices.

2) Triggering of the calculation of reconstructability characteristics of the incoming data upon local samples at the network edge (gateway) and choosing a data reduction mechanism based on these and the backend requirement.
3) Mode-based operation of the gateway which forwards, caches, or filters incoming data points to the backend system based on the previously chosen mechanism.

The present invention has inter alia the following advantages: The present invention enhances the efficiency in terms of storage, bandwidth and average data quality, preferably in an internet-of-thing system simultaneously maintaining the desired level for the reconstructability of the original data from the subset that has been stored in a data center.

Many modifications and other embodiments of the invention set forth herein will come to mind to the one skilled in the art to which the invention pertains having the benefit of the teachings presented in the foregoing description and the associated drawings. Therefore, it is to be understood that the invention is not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

Claims

1. A method for filtering data series (TS), preferably time series of data, prior to further processing, wherein the data series (TS) are collected by collecting entities (GW) and provided to one or more filtering entities (GW, DH) from one or more data delivering devices (D), and wherein the filtered data series (TS-RM) are forwarded to further processing entities (BS), characterized in that
   the filtering by said filtering entities (GW) is performed by the steps of
   a) Collecting a data series (O-TS),
   b) Reducing the information of the data series (O-TS) based on and by at least one data reduction procedures (RM1, RM2),
   c) Reconstruction of the original information (O-TS) for each reduced information of a data series (TS1-RM1, TS1-RM2),
   d) Calculating the level of reconstruction (RCL) for the information based on a comparison between the reconstructed information (TS1-RM1-RC1, TS1-RM2-RC1, TS1-RM2-RC2) and the original information (O-TS) collected in step a) for at least one of the reduction procedures (RM1, RM2), and
   e) Determining the reduced or non-reduced information of the data series (TS1, TS2) to be forwarded based on a comparison between a desired level and the calculated level (RCL) of reconstruction.

2. The method according to claim 1, characterized in that at least steps a)-d) are performed in irregular and/or regular time intervals, upon prespecified changes and/or appearance of prespecified values (V) in data series (TS1, TS2).

3. The method according to one of the claims 1-2, characterized in that when collecting the data series (TS1, TS2) the highest possible polling rate and/or the highest possible resolution is used.

4. The method according to one of the claims 1-3, characterized in that a reconstructability information (RT) is generated specifiying for each data series (TS1, TS2) and for each reduction procedure (RM1, RM2) and for corresponding input values for said reduction procedures (RM1, RM2) a value for the level of reconstruction (RCL).

5. The method according to claims 2 and 4, characterized in that the reconstructability information (RT) are updated when steps a)-d) are performed

6. The method according to one of the claims 1-5, characterized in that a reduction procedure (RM1, RM2) is provided in form of a procedure reducing dimensionality and/or size of the data series (TS1, TS2) and/or a generation of a function representing the data series (TS1, TS2).

7. The method according to one of the claims 1-6, characterized in that the comparison according to step d) is performed on a similarity metric, preferably using the Euclidian distance.

8. The method according to one of the claims 1-7, characterized in that the collecting entities (GW) are configured based on the operational status of said filtering entities (DH, GW).

9. The method according to claim 8, characterized in that when the operational status of the filtering entities (DH, GW) is dedicated for energy saving then the collecting entities (GW) are reconfigured such that only reduced information (TS1-RM1, TS1-RM2) satisfying the desired level of reconstruction is collected.

10. The method according to claim 8, characterized in that when the operational status of the filtering entities (DH, GW) is dedicated for network resource saving then the collecting entities (GW) are reconfigured such that only reduced information satisfying the desired level of reconstruction (DRCL) are forwarded and preferably the

collected information is cached in the filtering entity (DH) and/or in the collecting entity (GW).

11. The method according to claim 10, characterized in that when the operational status of the filtering entities (DH, GW) is dedicated for network resource saving, the collected information (O-TS) are forwarded upon demand of the further processing entities, in regular time intervals and/or never.

12. A system for filtering data series (TS), preferably time series of data, prior to further processing, comprising
one or more data delivering devices (D) adapted to provide data series (TS),
one or more collecting entities adapted to collect said data series and to provide them to one or more filtering entities (DH, GW) and wherein
said one or more filtering entities (GW, DH) are adapted to forward the filtered data series to further processing entities (BS), characterized in that
said one or more filtering entities (GW, DH) are adapted to perform the steps of
a) Collecting a data series (TS),
b) Reducing the information of the data series (O-TS) based on and by at least one data reduction procedures (RM1, RM2),
c) Reconstruction of the original information (O-TS) for each reduced information of a data series (TS1-RM1, TS2-RM2),
d) Calculating the level of reconstruction (RCL) for the information based on a comparison between the reconstructed information (TS1-RM1-RC, TS1-RM2-RC) and the original information (O-TS) collected in step a) for at least one of the reduction procedures (RM1, RM2), and
e) Determining the reduced or non-reduced information of the data series (TS1, TS2) to be forwarded based on a comparison between a desired level (DRCL) and the calculated level (RCL) of reconstruction.

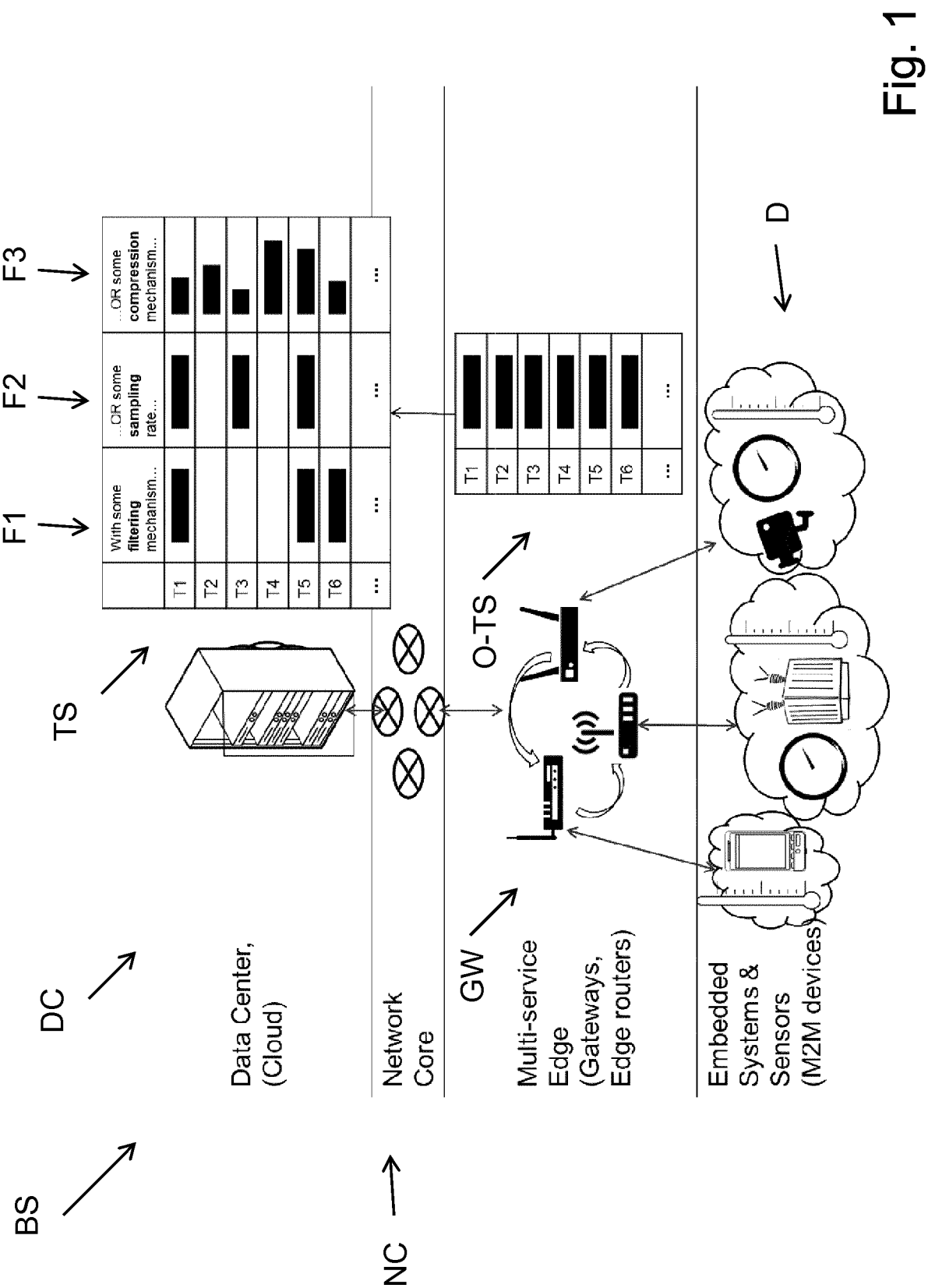

BS
DC
NC
GW
TS
O-TS
D
F1
F2
F3
With some filtering mechanism...
...OR some sampling rate...
...OR some compression mechanism...
T1
T2
T3
T4
T5
T6
Data Center, (Cloud)
Network Core
Multi-service Edge (Gateways, Edge routers)
Embedded Systems & Sensors (M2M devices)

Fig. 1

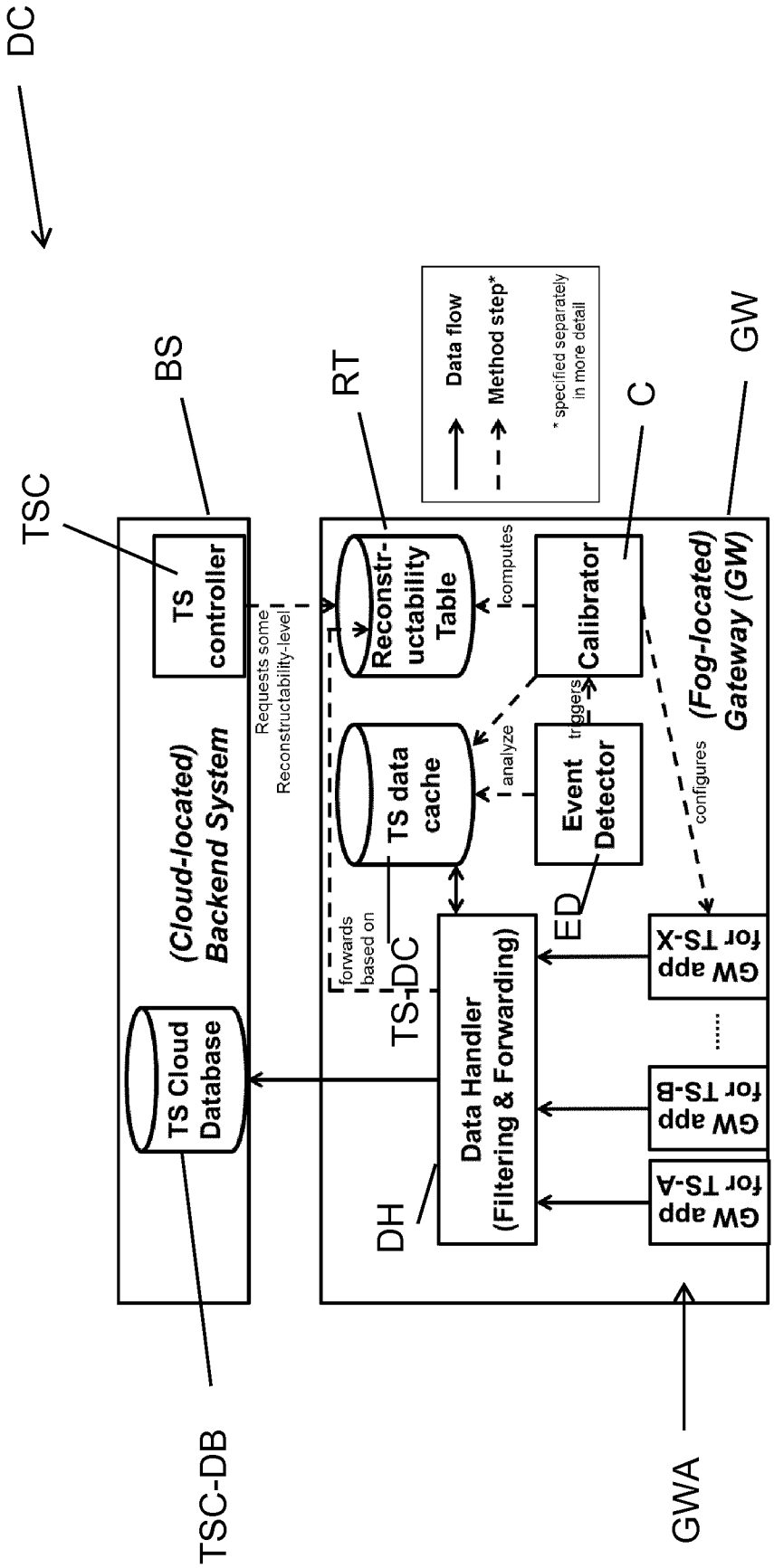

DC
TSC
BS
RT
C
GW
Data flow
Method step*
* specified separately in more detail
TS controller
Reconstr-uctability Table
computes
Calibrator
(Cloud-located) Backend System
Requests some Reconstructability-level
TS data cache
analyze
Event Detector
triggers
configures
(Fog-located) Gateway (GW)
forwards based on
TS-DC
ED
TS Cloud Database
Data Handler (Filtering & Forwarding)
GW app for TS-X
.....
GW app for TS-B
GW app for TS-A
DH
TSC-DB
GWA

Fig. 2

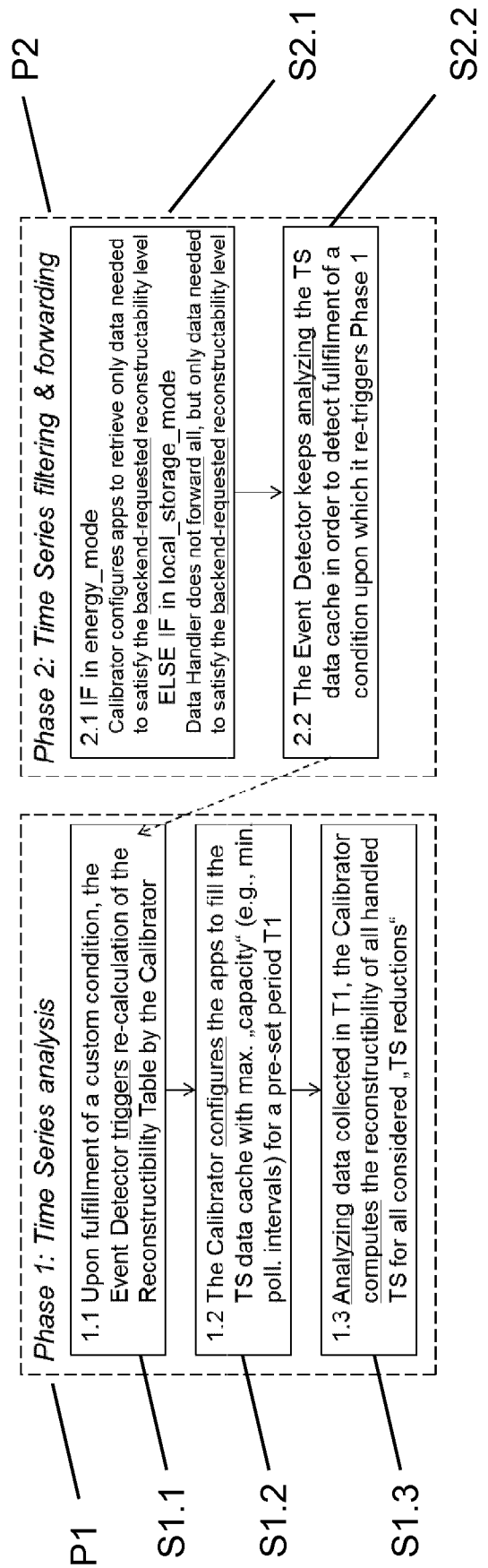

P1
S1.1
S1.2
S1.3
P2
S2.1
S2.2
Phase 1: Time Series analysis
1.1 Upon fulfillment of a custom condition, the Event Detector triggers re-calculation of the Reconstructibility Table by the Calibrator
1.2 The Calibrator configures the apps to fill the TS data cache with max. „capacity“ (e.g., min poll. intervals) for a pre-set period T1
1.3 Analyzing data collected in T1, the Calibrator computes the reconstructibility of all handled TS for all considered „TS reductions“
Phase 2: Time Series filtering & forwarding
2.1 IF in energy_mode
Calibrator configures apps to retrieve only data needed to satisfy the backend-requested reconstructability level
ELSE IF in local_storage_mode
Data Handler does not forward all, but only data needed to satisfy the backend-requested reconstructability level
2.2 The Event Detector keeps analyzing the TS data cache in order to detect fullfilment of a condition upon which it re-triggers Phase 1

Fig. 3

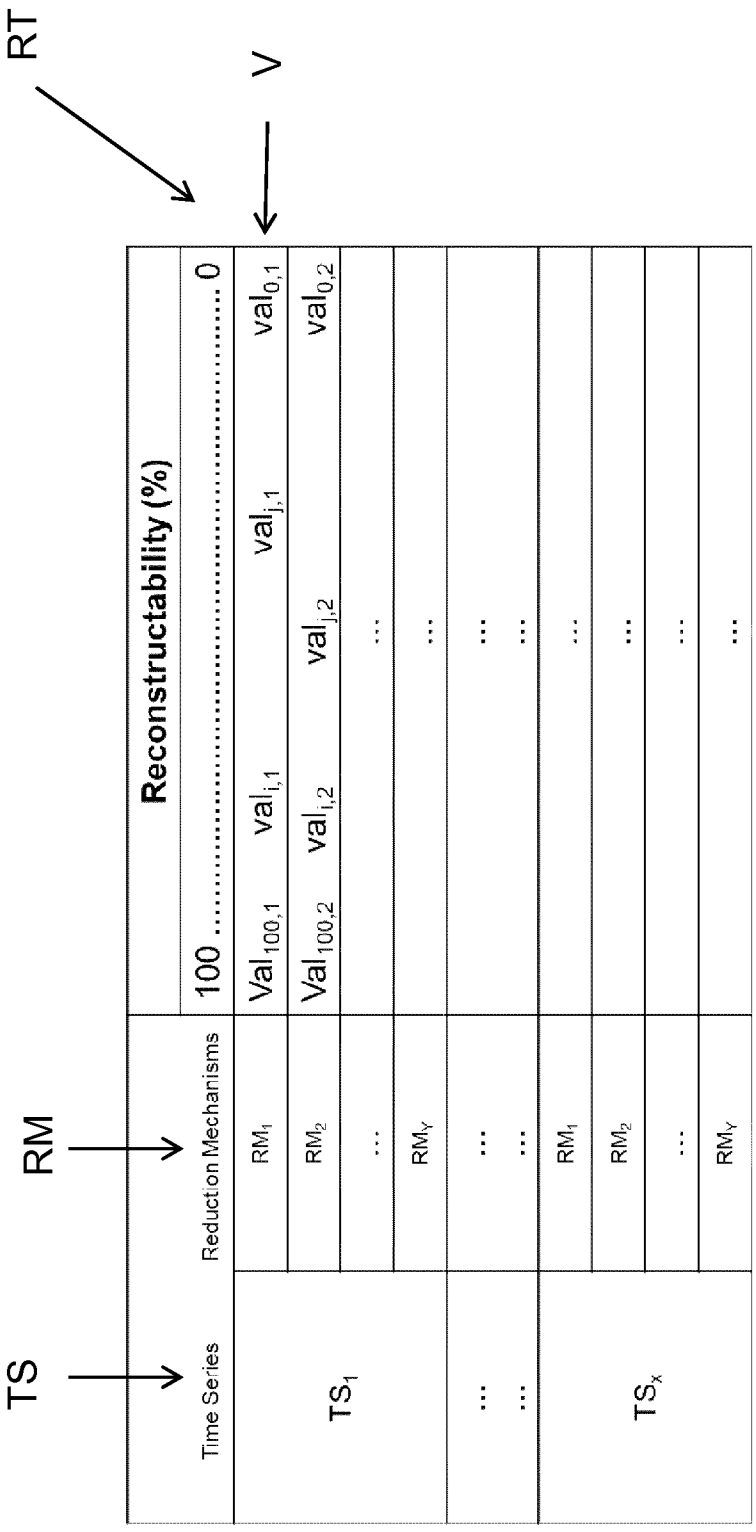

RT
V
TS
RM
Time Series
Reduction Mechanisms
Reconstructability (%)
100 0
TS1
TSx
RM1
RM2
RMY
Val100,1 vali,1 valj,1 val0,1
Val100,2 vali,2 valj,2 val0,2

Fig. 4

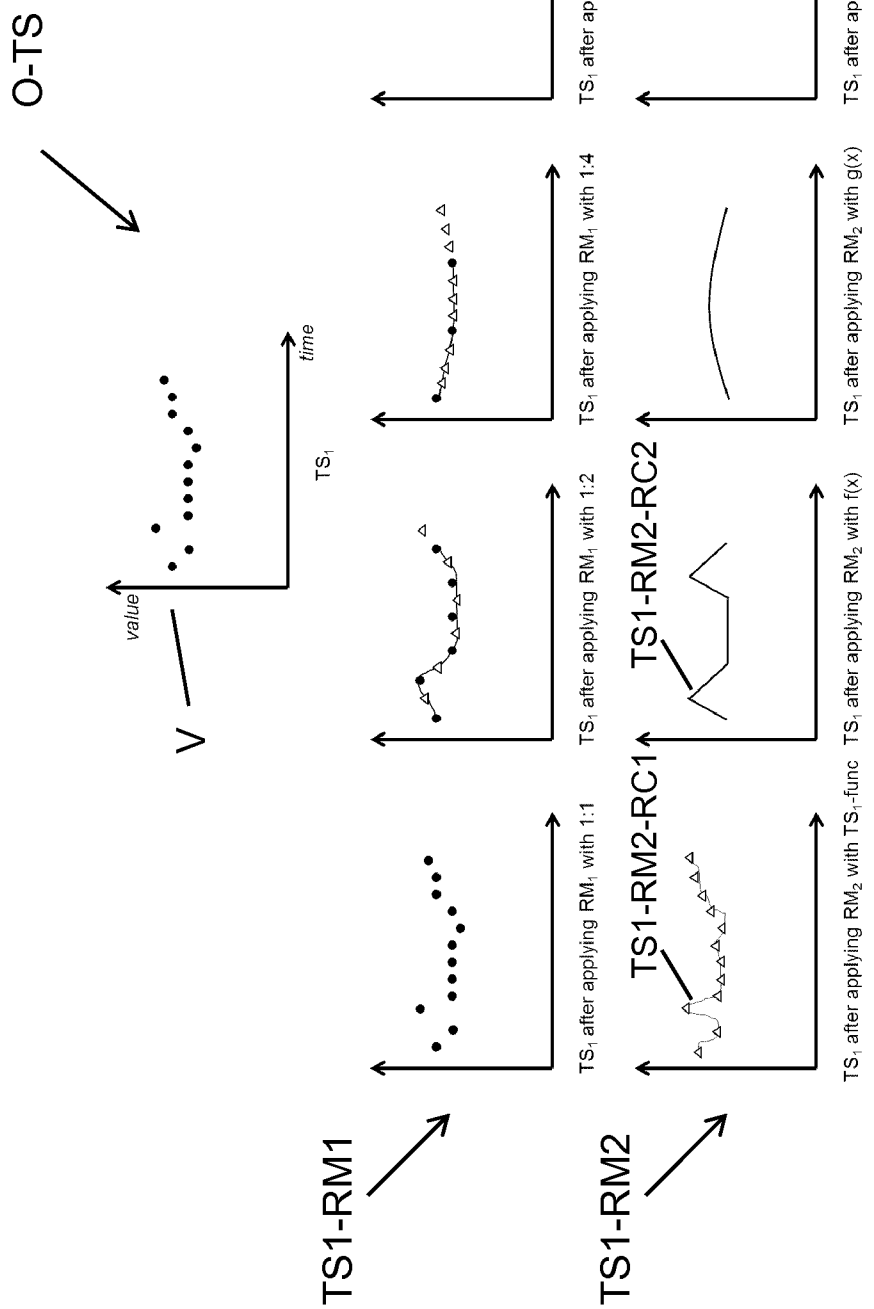

O-TS
V
value
time
TS1
TS1-RM1
TS1-RM2
TS1-RM2-RC1
TS1-RM2-RC2
TS1 after applying RM1 with 1:1
TS1 after applying RM1 with 1:2
TS1 after applying RM1 with 1:4
TS1 after applying RM1 with 0:∞
TS1 after applying RM2 with TS1-func
TS1 after applying RM2 with f(x)
TS1 after applying RM2 with g(x)
TS1 after applying RM2 with 0:∞

Fig. 5

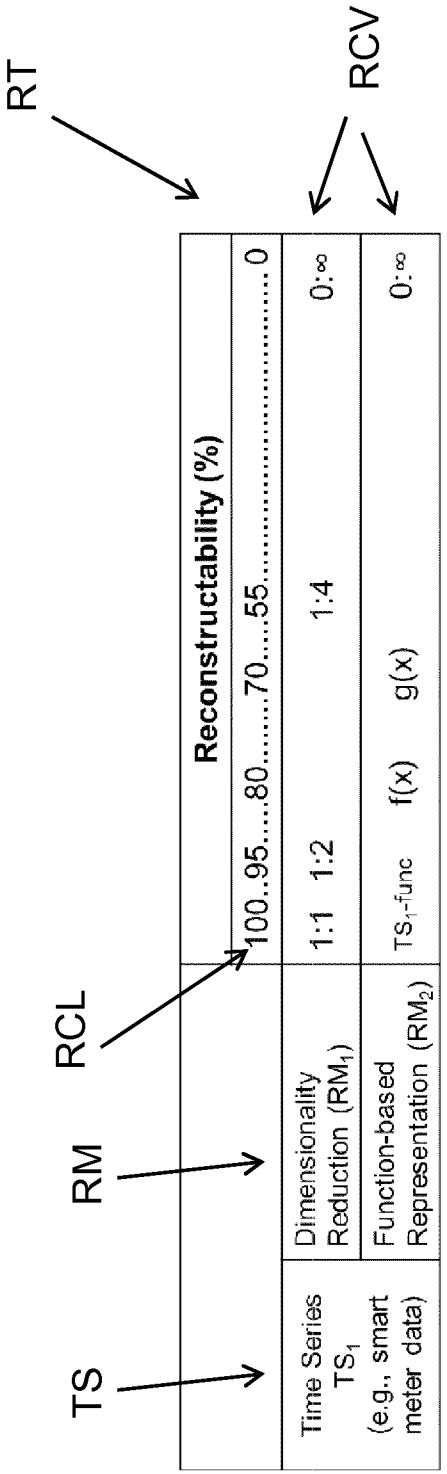

RT
RCV
RCL
RM
TS
Reconstructability (%)
100..95.....80.....70.....55.....0
Time Series TS_1 (e.g., smart meter data)
Dimensionality Reduction (RM_1)
1:1 1:2 1:4 0:∞
Function-based Representation (RM_2)
TS_1-func f(x) g(x) 0:∞

Fig. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/076899

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04L29/08 H04W4/00
ADD.

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)
H04L H04W

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2009/003599 A1 (HART BRETT D [US] ET AL) 1 January 2009 (2009-01-01) abstract figure 1 paragraph [0012] - paragraph [0025] -----	1-12
X	US 2014/359035 A1 (WANG CHONGGANG [US] ET AL) 4 December 2014 (2014-12-04) abstract figure 1 paragraph [0026] - paragraph [0058] -----	1-12

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

☒ See patent family annex.

* 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
2 June 2015	10/06/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Canosa Areste, C

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2014/076899

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009003599 A1	01-01-2009	EP 2009876 A2	31-12-2008
		US 2009003599 A1	01-01-2009
US 2014359035 A1	04-12-2014	US 2014359035 A1	04-12-2014
		WO 2014193950 A1	04-12-2014