

(19)



SUOMI - FINLAND

(FI)

PATENTTI- JA REKISTERIHALLITUS

PATENT- OCH REGISTERSTYRELSEN

FINNISH PATENT AND REGISTRATION OFFICE

(10) FI 130073 B

(12) PATENTTIJULKAISU

PATENTSKRIFT

PATENT SPECIFICATION

(45) Patentti myönnetty - Patent beviljats - Patent granted

31.01.2023

(51) Kansainvälinen patenttiluokitus - Internationell patentklassifikation - International patent classification

H04L 12/28 (2006.01)

H04L 41/149 (2022.01)

H04L 41/16 (2022.01)

H04L 43/0817 (2022.01)

G06N 20/00 (2019.01)

G05B 23/02 (2006.01)

(21) Patenttihakemus - Patentansökning - Patent application

20216107

(22) Tekemispäivä - Ingivningsdag - Filing date

27.10.2021

(23) Saapumispäivä - Ankomstdag - Reception date

27.10.2021

(43) Tullut julkiseksi - Blivit offentlig - Available to the public

31.01.2023

(73) Haltija - Innehavare - Proprietor
1 • **Elisa Oyj**, Ratavartijankatu 5, 00520 Helsinki, SUOMI - FINLAND, (FI)(72) Keksijä - Uppfinnare - Inventor
1 • **NIKITIN, Alexander**, HELSINKI, SUOMI - FINLAND, (FI)(74) Asiamies - Ombud - Agent
Espatent Oy, Kaivokatu 10 D, 00100 Helsinki(54) Keksinnön nimitys - Uppfinningens benämning - Title of the invention
Prediktiivinen kaapelimodeemien kunnossapito
Prediktivt underhåll av kabelmodem
Predictive maintenance of cable modems(56) Viitejulkaisut - Anförda publikationer - References cited
US 10637715 B1, US 2021250325 A1, US 2019258904 A1, US 2020322662 A1, CN 113076694 A

(57) Tiivistelmä - Sammandrag - Abstract

Menetelmä, tietokoneohjelma ja laite kaapelimodeemiterminointijärjestelmän CMTS ennakoivaan ylläpitoon, jossa automaattisesti: jatkuvasti kerätään (710) kaapelimodeemiterminointijärjestelmän tilatietoja; rakennetaan (720) koneoppimisaineisto kerätystä tiedoista, sisältäen sen, että: kerätään (730) aiempia havaintotietoja tietojaksolta kerätystä tilatiedoista; ennustetaan (740) tulevia havaintotietoja ennustejakson aikana käyttäen luokitustehtävää, että onko tietyssä laitteessa ongelma ennustejakson aikana ja regressiotehtävää ennustamaan ajan seuraavaan ongelmalliseen tilanteeseen; käytetään (750) koneoppimistietojoukkoa koneoppimismallilla, joka on valittu joukosta, jonka muodostavat gradienttitehostus, äärisatunnaistetut puut; ja hermoverkot; ja sisällytetään (760) kaapelimodeemiterminointijärjestelmän paikkatietoa; ja lisäksi käytetään (770) koneoppimista rakennetun koneoppimisen tietojoukon kanssa sellaisten laitteiden ennustamiseksi, jotka ovat alttiita ylläpito-ongelmille ennustejakson aikana.

A method, computer program and apparatus for predictive cable modem termination system, CMTS, maintenance including automatically: continually collecting (710) state data of a CMTS; constructing (720) a machine learning, ML, dataset from the collected data, including: accruing (730) past observation data from an information period from the collected state data; predicting (740) future observation data during a prediction period using a classification task will there be a problem with a particular piece of equipment during the prediction period and a regression task to predict time until a next problematic situation; using (750) the ML dataset by an ML model selected from a group of gradient boosting; extremely randomised trees; and neural networks; and incorporating (760) spatial knowledge of the CMTS; and further using (770) the ML with the constructed ML dataset to predict pieces of equipment prone to experience maintenance issues during the prediction period.

710. continually collecting state data of a cable modem termination system
720. constructing a machine learning dataset from the collected data, including:
730. accruing past observation data from an information period from the collected state data
740. predicting future observation data during a prediction period using a classification task will there be a problem with a particular piece of equipment during the prediction period and a regression task to predict time until a next problematic situation
750. using the machine learning dataset by a machine learning model selected from a group of gradient boosting; extremely randomised trees; and neural networks; and including
760. incorporating spatial knowledge of the cable modem termination system
770. using the machine learning with the constructed machine learning dataset to predict pieces of equipment prone to experience maintenance issues during the prediction period

PREDICTIVE MAINTENANCE OF CABLE MODEMS

TECHNICAL FIELD

The present disclosure generally relates to predictive maintenance of cable modems. The disclosure relates particularly, though not exclusively, to training and using of machine learning for the predictive maintenance of the cable modems.

BACKGROUND

This section illustrates useful background information without admission of any technique described herein representative of the state of the art.

A cable TV network includes a cable modem termination system (CMTS), several amplifiers, and a plurality of end-user cable modems. The data are electrically and / or optically transported in two directions, upstream and downstream, in different frequencies. The connections on each leg can be measured by each recipient. This enables collecting from the CMTS both temporal and spatial data concerning different transmitters. Moreover, internal state and diagnostic data can also be collected and used for comparing with the measurement data.

As a CMTS typically contains a plurality of parallel smaller equipment clusters, aforementioned data could be collected from a number of theoretically similar sets of equipment. However, settings, hardware and firmware tend to vary, so also on a cluster level, the collected data varies from one cluster to another. Hence, the data collected in different clusters cannot be simply compared to identify maintenance issues. Moreover, the settings and states of the equipment vary through fluctuation of use as well as changes in settings and condition of different pieces of equipment.

SUMMARY

It is an object of the present invention to provide for predictive maintenance of cable modems. Another or alternative object is to efficiently train machine learning for predictive maintenance of cable modems. Yet another or alternative object is to efficiently use the machine learning for predictive maintenance of cable modems.

The appended claims define the scope of protection. Any examples and technical descriptions of apparatuses, products and / or methods in the description and / or drawings not covered by the claims are presented not as embodiments of the invention but as background art or examples useful for understanding the invention.

According to a first example aspect there is provided a method for predictive cable modem termination system maintenance, comprising automatically:

continually collecting state data of a cable modem termination system;

constructing a machine learning dataset from the collected data, including:

5 accruing past observation data from an information period from the collected state data;

 predicting future observation data during a prediction period using a classification task will there be a problem with a particular piece of equipment during the prediction period and a regression task to predict time until a next problematic situation;

10 using the machine learning dataset by a machine learning model selected from a group of gradient boosting; extremely randomised trees; and neural networks; and

 incorporating spatial knowledge of the cable modem termination system;

 the method further comprising using the machine learning with the constructed machine learning dataset to predict pieces of equipment prone to experience maintenance issues during the prediction period.

20 The predicting of the pieces of equipment prone to experience maintenance issues during the prediction period may comprise determining a set of most likely pieces of equipment to experience the maintenance issues. The set may have a constant number of pieces of equipment. The predictions of the most likely pieces of equipment to experience the maintenance issues may be referred to predictions.

25 The predicting of the pieces of equipment prone to experience maintenance issues during the prediction period may comprise determining a precision of the predicting by a ratio of a number of correct positives divided by a number of all predicted pieces of equipment. The predicting of the pieces of equipment prone to experience maintenance issues during the prediction period may comprise adapting the given number of most likely pieces of equipment according to development of the ratio. The given number may be increased by N_i per cent if the ratio exceeds a threshold over a given adaptation period. The adaptation period may be at least 5 days; at least 7 days; at least 10 days; at least 14 days; or at least 30 days. The adaptation period may be at most 7 days; at most 10 days; at most 14 days; at most 30 days; or at most 60 days. N_i may be at least 1; 2; 5; 10; 15; or 20. N_i may be at most 2; 5; 10; 15; 20; or 50. N_i may be proportional or inversely proportional to a difference of the ratio from the threshold.

35 The predicting of the pieces of equipment prone to experience maintenance issues during the prediction period may comprise choosing the machine learning model from two or more alternatives using a cross-validation. The cross-validation may be performed dividing the

data set into at least 2; 3; 4; 5; 6; or 8 folds. The cross-validation may be performed with a mean absolute error metric. After performing the cross-validation, gradient boosting may be performed, such as using Friedman's Greedy function approximation. In the gradient boosting, following values may be used: min child weight: 1 to 10; gamma: 0.5 to 5; subsample: 0.6 to 1.0, colsample bytree: 0.6 to 1.0; max depth: 3 to 7; n estimators: 100 to 500.

Continually collecting may refer to collecting with regular intervals or periodically. Alternatively, the continually collecting may refer to repeatedly collecting with regular or irregular intervals. The continually collecting may refer to collecting as a continuous process over days, weeks, or years, when predictive maintenance of cable modems is provided.

The state data may be collected as Internet Protocol Detailed Records, IPDRs. The IPDRs may comply with the Broadband Forum TR-232 Issue 1 of May 2012.

The state data may comprise for one or more amplifiers and / or other network elements in a cable modem system: statuses; connectivity information; geographical location information; diagnostic warnings; and / or system messages. The state data may further comprise information about connectivity of the cable modem termination network, such as information about which pairs of the devices are connected; events that occurred with devices in the cable modem termination system, such as warnings or bad connectivity notices; severity; and / or one or more timestamps.

The state data may further comprise service tickets issued by user as an indicator of breaks or other issues in the cable modem termination network.

The pieces of equipment may comprise one or more amplifiers. The pieces of equipment may comprise one or more cable modems.

The method may further comprise automatically taking corrective action to mitigate the predicted maintenance issues.

The state data may comprise one or more of following: a signal to noise ratio, such as an Internet Protocol Detailed Record field CmtsCmUsSignalNoise; a number of corrected codewords, such as an Internet Protocol Detailed Record field CmtsCmUsCorrecteds; a cable upstream equalization-coefficient, such as an Internet Protocol Detailed Record field CmtsCmUsEqData; an upstream high resolution timing offset which represents a round trip time on an upstream channel of a cable modem, e.g., in units of $(6.25 \text{ microseconds} / (64 * 256))$, such as an Internet Protocol Detailed Record field CmtsCmUsHighResolutionTimingOffset, wherein this value may be set to zero when the

measurement is unknown; a Boolean flag for muting state of the channel, optionally marked as true if the upstream channel of the cable modem has been muted, e.g., by a CM-CTRL-REQ/CM-CTRL-RSP message exchange, such as an Internet Protocol Detailed Record field CmtsCmUsIsMuted; a number of micro reflections received on an interface, such as an Internet Protocol Detailed Record field CmtsCmUsMicroreflections; a modulation type used by a given channel, optionally with corresponding values: 0: unknown, 1: tdma, 2: atdma, 3: scdma, 4: tdmaAndAtdma, such as an Internet Protocol Detailed Record field CmtsCmUsModulationType; a ranging status of a cable modem, optionally with corresponding values: 1:other, 2:aborted, 3:retriesExceeded, 4:success, 5:continue, such as an Internet Protocol Detailed Record field CmtsCmUsRangingStatus; a receive power as perceived for the upstream channel, such as an Internet Protocol Detailed Record field CmtsCmUsRxPower; a number of codewords received with uncorrectable errors from the cable modem, such as an Internet Protocol Detailed Record field CmtsCmUsUncorrectables; a number of codewords received without any errors, such as an Internet Protocol Detailed Record field CmtsCmUsUnerrored); and / or a system uptime value taken from CMTS when the IPDR record was created, such as an Internet Protocol Detailed Record field CmtsSysUpTime).

The information period 310 may comprise a plurality of considering periods such that subsequent considering periods represent successive intervals of time.

A first group of features may be extracted from the Internet Protocol Detailed Records. The first group of features may comprise N_1 elements representing minima over the information period.

A second group of features may be extracted from the Internet Protocol Detailed Records. The second group of features may comprise N_2 elements representing maxima over the information period. N_2 may equal to N_1 .

A third group of features may be extracted from the Internet Protocol Detailed Records. The third group of features may comprise N_3 elements representing averages over the information period. N_3 may equal to N_1 .

A fourth group of features may be extracted from the Internet Protocol Detailed Records. The fourth group of features may comprise N_4 elements representing latest values over the information period, that is, values of the latest considering period. N_4 may equal to N_1 .

A plurality of upstream channels may be represented by a joint average. Alternatively, features representing different upstream channels may be combined by concatenation.

The collecting of the state data may comprise pulling data from the cable modem termination system in an extended markup language format. The collecting of the state data may comprise providing the pulled data in the extended markup language format to a distributed event streaming platform such as Kafka. The pulled data may be subsequently stored in a cloud storage. The cloud storage may be S3 compatible. The storing into the cloud storage may be implemented using Ceph. The collected data may comprise a plurality of Internet Protocol Detailed Record schemas, comprising one or more of: upstream-util stats, such as CMTS-US-UTIL, downstream-util stats, such as CMTS-DS-UTIL, topology type, such as CMTS-TOPOLOGY-TYPE, upstream type, such as CMTS-US-UTIL-STATSTYPE, registration status, such as CMTS-CM-REG-STATUS-TYPE, and quality of service, such as DOCSIS-QOS.

The collecting of the state data may be performed by a controller.

The training of the machine learning model may be performed using a binary classification task and a regression task.

The incorporating of the spatial knowledge of the cable modem termination system may comprise forming a separate predictor of issues in the cable modem termination system. The forming of the separate predictor may comprise generating numerical embeddings of each vertex in a network graph, passing these representations to the machine learning model as additional inputs, and using the model.

The vertex embeddings may be computed by taking corresponding rows from an adjacency matrix.

Alternatively, other approaches may be used to generate a meaningful representation of the graph vertices, such as node2vec or different modifications of biased random walks.

If the system works on top of aggregated data, the data may be aggregated via all the intervals with minimum, maximum, average, and percentile operations. Automatic feature extraction techniques may be performed in this step to improve predictive performance.

In an alternative embodiment, a joint predictor is computed using both network information and temporal information from the collected state data.

Recently repaired cable modems may be placed in a quarantine or a state in which corrective action is not taken on issues unless they persist for a time exceeding a given threshold. Otherwise, the method may comprise automatically triggering a proactive service ticket for predicted issues for corrective action.

If it is found on taking corrective action that the predicted issue was correctly predicted, such an observation may be propagated to the machine learning model.

In an embodiment, the method is deployed using computer cloud technologies. Source code may be stored in a code repository for different services. When the system is deployed, the code may be delivered to a Jenkins server. Docker containers may be built with the code. Built docker containers may be run as Airflow jobs. The predictions may be stored in Splunk.

The machine learning models may be retrained by API requests. The machine learning models may be stored in an S3-compatible storage. A newest model may be retrieved by an inference server and used for periodic, such as daily, predictions. Alternatively, and / or additionally, without any additional development, the system may be used more frequently than once a day, even as a real-time tool. Alternatively, and / or additionally, without any additional development, the system may be used less frequently than once a day, for example, by averaging daily predictions.

Different machine learning models may be formed and separated by cable modem termination system, if accommodate significant differences in measurements and functioning. Alternatively, one machine learning model may be used for two or more cable modem termination systems. In an example embodiment, a difference between two or more cable modem termination systems is measured so that if there are enough data collected from all the cable modem termination systems for training separate models, then machine learning models are trained separately and a singular model is trained for all (or a subset) of the cable modem termination systems, and the performance of individual and singular machine learning models is tested on a hold-out dataset or by backtesting. Suitability of a singular model may be further verified using statistical tests to select a suitable model or models. In another example embodiment, in which there are insufficient data collected, the method comprises transforming collected state data in a lower dimensional space (for example, using principle component analysis (PCA)); performing a multiple sample testing, for instance with maximum mean discrepancy (MMD); and obtaining p-values (for example, using permutations test on a kernel matrix extracted from MMD kernel matrix), indicative of which cable modem termination systems are sufficiently similar to allow joint use of a same machine learning model.

The cable modem termination system may comprise more than 1000 cable modems. The cable modem termination system may comprise more than 10000 cable modems.

According to a second example aspect there is provided a computer program comprising computer executable program code which when executed by at least one processor causes

an apparatus at least to perform the method of the first example aspect.

According to a third example aspect there is provided an apparatus comprising means for performing the method of the first example aspect. The means may comprise the computer program of the second example aspect. The means may comprise at least one processor
5 configured to execute the program code.

According to a fourth example aspect there is provided a computer program product comprising a non-transitory computer readable medium having the computer program of the third example aspect stored thereon.

Any foregoing memory medium may comprise a digital data storage such as a data disc or
10 diskette; optical storage; magnetic storage; holographic storage; opto-magnetic storage; phase-change memory; resistive random-access memory; magnetic random-access memory; solid-electrolyte memory; ferroelectric random-access memory; organic memory; or polymer memory. The memory medium may be formed into a device without other substantial functions than storing memory or it may be formed as part of a device with other
15 functions, including but not limited to a memory of a computer; a chip set; and a sub assembly of an electronic device.

Different non-binding example aspects and embodiments have been illustrated in the foregoing. The embodiments in the foregoing are used merely to explain selected aspects or steps that may be utilized in different implementations. Some embodiments may be
20 presented only with reference to certain example aspects. It should be appreciated that corresponding embodiments may apply to other example aspects as well.

BRIEF DESCRIPTION OF THE FIGURES

Some example embodiments will be described with reference to the accompanying figures, in which:

- 25 Fig. 1 schematically shows a system according to an example embodiment;
- Fig. 2 schematically shows an environment in which some computation tasks are performed in an example embodiment;
- Fig. 3 schematically illustrates data collecting and target representation from raw data;
- Fig. 4 shows an example recurrent neural network, RNN, architecture for prediction of time
30 until the next malfunction;
- Fig. 5 shows an exemplary node graph of a cable network;
- Fig. 6 shows a block diagram of an apparatus according to an example embodiment; and
- Fig. 7 shows a flow chart according to an example embodiment.

DETAILED DESCRIPTION

In the following description, like reference signs denote like elements or steps.

Various embodiments are next described for predictive cable modem termination system maintenance in a cable modem network. Cable modems may be connected to amplifiers and then connected to the cable modem termination system.

Fig. 1 schematically shows a system 100 according to an example embodiment for predictive cable modem termination system maintenance. In Fig. 1, arrows indicate transfer of state data and predicted service tickets rather than user data delivered via cable modems.

The system 100 comprises a plurality of cable modems, of which Fig. 1 presents a first cable modem 110 that is functioning normally and a second cable modem 110' that is experiencing a problem or issue. The system 100 further comprises one or more amplifiers 120 and a data collector 130 for collecting data 112 from the cable modems and from the amplifiers 122. The amplifiers may also obtain data 114, 114' from the cable modems 110, 110'. The data collector 130 stores 132 collected data to a database 140.

Service tickets may be issued by pieces of equipment or their users 116 to a ticketing system 150. The ticketing system stores 152 the service tickets to the database 140.

A predictor 160 reads 142 from the database 140 a machine learning model and recent data and adds 162 predicted tickets to the ticketing system 150.

The system 100 further comprises machine learning equipment 170 for training or updating the machine learning model. The machine learning equipment 170 reads 144 the collected data from the database 140 and writes 172 a new version of the machine learning model to the database 140.

Fig. 2 schematically shows an environment in which some computation tasks are performed in an example embodiment. Fig. 2 shows a controller 210; a Kafka functionality 220; and a storage 230 connected to the Internet 240. In an example embodiment, data collector 130 operates as the controller 210 or the controller 210 further operates as the data collector 130.

In an example embodiment, predictive modelling of the cable modem termination system is performed in the following steps:

- data collection,
- data preparation,
- training of the models,
- running API,

- performing maintenance actions.

These steps will be described in further detail in the following.

Data collection

This step's primary goal is to collect the data that describes the condition of the cable network well. For this purpose, Internet Protocol Detailed Record measurements are collected, for example, every 15 minutes of the snapshots of the data from cable modem termination systems. A complete list of measurements collected can be found at <http://mibs.cablelabs.com/namespaces/DOCSIS/3.0/xsd/ipdr/>. In an example embodiment, the features extracted from upstream and downstream data were used; although, it can be used other classes of data from Internet Protocol Detailed Record (e.g., topology data and registration status). Furthermore, information was collected about network topology and measurements from amplifiers and cable modem termination systems.

Information was also collected about the statuses of amplifiers and other network elements: connectivity and geographical location of amplifiers, warnings, and other system messages that happened in amplifiers. The additional data contains information about the network's connectivity (which pairs of the devices are connected), events that occurred with the device (e.g., warning, or bad connectivity notice), severity, timestamp.

Still further, data was collected about the users' problem reports used as an indicator of the breaks.

Data preparation

In this step, a machine learning dataset is constructed from the collected data. Conversion of raw IPDR (internet protocol detail record) data to the machine learning dataset is schematically shown in Fig. 3.

Here, Internet Protocol Detailed Record data is used to represent a state of a cable modem state and service tickets are used as an indicator of a problem. Tickets are the records of the problems created in our inner system following problem reports of cable modem users. Fig. 3 schematically illustrates data collecting and target representation from raw data. An information period 310 refers to a time period used for extracting feature representation of a prediction time point. At this step, we have temporal measurements from the cable modem termination systems. In one example embodiment, aggregation (maximum, minimum, average, and percentiles) is applied to these time-series data. In another example embodiment, temporal models are applied, such as a long short-term memory, a gated recurrent unit, and other types of recurrent neural networks.

From the Internet Protocol Detailed Record data, following features were found informative about the condition of a cable modem:

1. Signal to noise ratio (IPDR field: CmtsCmUsSignalNoise),
2. Number of corrected codewords (IPDR field: CmtsCmUsCorrecteds),
- 5 3. The cable upstream equalization-coefficient (IPDR field: CmtsCmUsEq- Data),
4. Upstream high resolution timing offset which represents the round trip time on this CM's upstream channel in units of (6.25 microseconds/(64*256)). This value is set to zero when the measurement is unknown (IPDR field: CmtsCmUsHighResolutionTimingOffset),
- 10 5. Boolean flag is the channel is muted. Marked as true if the CM's upstream channel has been muted via CM-CTRL-REQ/CM-CTRL-RSP message exchange(IPDR field: CmtsCmUsIsMuted),
6. The number of microreflections received on this interface (IPDR field: CmtsCmUsMicroreflections),
- 15 7. The modulation type used by a given channel, with corresponding values: 0: unknown, 1: tdma, 2: atdma, 3: scdma, 4: tdmaAndAtdma (IPDR field: CmtsCmUsModulationType),
8. Ranging status of the cable modem, with corresponding values: 1:other, 2:aborted, 3:retriesExceeded, 4:success, 5:continue (IPDR field: CmtsC- mUsRangingStatus),
- 20 9. The receive power as perceived for the upstream channel (IPDR field: CmtsCmUsRxPower),
10. The number of codewords received with uncorrectable errors from the CM (IPDR field: CmtsCmUsUncorrectables),
11. The number of codewords received without any errors (IPDR field: CmtsCmUsUnerroreds),
- 25 12. System uptime value taken from CMTS when the IPDR record was created(IPDR field: CmtsSysUpTime).

Further information about the aforementioned fields is available in Data-Over-Cable Service Interface Specifications DOCSIS 3.0 by Cable Television Labs Inc, available at <https://www.cablelabs.com/wp-content/uploads/2015/08/CM-SP-OSSlv3.0-I05-071206.pdf>.

An example of extracted features: [299.0, 86.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, -21.0, 52.0, 1280836.0, 353.0, 9946.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, -12.0, 156.0, 2812610.0, 337.4878, 2809.9512, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, -18.4146, 86.6341, 1872500.4634, 336.0, 177.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, -20.0, 53.0, 1967139.0]

First 18 elements or extracted features are minima over the information period, the second 18 elements are maxima over the information period, the third 18 elements are averages over the information period, and the last (fourth) 18 elements are the last over information period, i.e., from the last considering period. Every 18 elements (of the listed above) are:

5 signal to noise ratio, number of corrected codewords, upstream high-resolution timing offset, normalized number of muted upstream channels, number of upstream micro reflections, upstream modulation type (five binary variables because one-hot encoding of the categorical variable), upstream ranging status (five binary variables because one-hot encoding of the categorical variable), receive power for the upstream channel, the number

10 of uncorrected codewords, the number of codewords received without any errors. In this example we calculated average of the features over all upstream channels from four to twelve (or, e.g., four to six), also these features can be extracted for each upstream channel and concatenated which allows to produce a more precise machine learning model.

In an example embodiment, the data collection comprises any one or more of following

15 steps: pulling the collected data from the cable modem termination systems in an extended markup language, XML, format, then the collected data are supplied to Kafka, and then using Kafka Connect to an S3 compatible storage, which in an example embodiment is implemented via Ceph. The collected data contains in an example embodiment the following Internet Protocol Detailed Record schemas: upstream-util stats (CMTS-US-UTIL),

20 downstream-util stats (CMTS-DS-UTIL), topology type (CMTS-TOPOLOGY-TYPE), upstream type (CMTS-US-UTIL-STATS-TYPE), registration status (CMTS-CM-REG-STATUS-TYPE), and quality of service (DOCSIS-QOS).

Training of the machine learning model

The dataset that is described in the foregoing is well suited to various machine learning

25 algorithms, including binary classification and regression tasks.

Classification task

In the classification task, two classes are considered: (i) normal functioning of a cable modem, and (ii) problematic situation of the cable modem. The classification problem can be described by a question: "Will there be a problem with the particular cable modem in

30 next T hours." Having this, we considered prediction time points in the dataset and collected Internet Protocol Detailed Record data over information period (W) similarly to Fig. 3, and consider period of time T after the prediction time point, wherein $1 \text{ h} \leq W \leq 48 \text{ h}$, $6 \leq T \leq 120 \text{ h}$. In an example embodiment, W is 4 h, 5 h, 6 h, 8 h, or 10 h. In an example embodiment, T is 20 h, or 25 h, or 30 h, or 35 h, or 40 h, or 45 h, or 50 h, or 60 h, or 72 h.

Regression task

We can also describe the problem of predicting future problems as a regression task. Instead of classifying the prediction time points, we predict time until the next problematic situation (marked as a target variable for regression in Fig. 3). We also identify the window sizes for the maximum future window (if a ticket was registered in time larger than this interval, we set a target variable equal to this time interval). In an example embodiment, identified parameters were $W = 6$ h and $T = 40$ h.

Methods

Experimentally, the regression task described in the foregoing was chosen in sake of a flexible target variable and empirical evaluations. However, other approaches may also work for those. Prior to the application of the machine learning methods, the data may be scaled with min-max scaling.

We have tested different machine learning models here, including gradient boosting by Friedman, 2001, extremely randomized trees by Geurts et al., 2006, neural networks by McCulloch and Pitts, 1943. A most suitable model (at least among tested alternatives) can be chosen, e.g., using cross-validation techniques with the dataset divided into five folds. After performing cross-validation with mean absolute error metric, a gradient boosting approach may be used, such as that by Friedman, 2001, with the following parameters: 'colsample bytree': 0.8, 'gamma': 0.5, 'max depth': 3, 'min child weight': 1, 'subsample': 1.0. The hyperparameters can vary for different installations and client but still can be found with the same techniques. In an example embodiment, the parameters were chosen from the following ranges: min child weight: 1 to 10, gamma: 0.5 to 5, subsample: 0.6 to 1.0, colsample bytree: 0.6 to 1.0, max depth: 3 to 7, n estimators: 100 to 500.

Neural network architectures were found to deliver good results, although they were also found to be less interpretable. For example, we also experimented with neural networks such as Long Short-Term Memory (LSTM) as a machine learning algorithm, and the performance was comparable to the chosen method. Fig. 4 shows an example Recurrent Neural Network, RNN, architecture for prediction of time until the next malfunction.

These architectures can be extended by the addition of convolutional, attention, recurrent layers in various directions, or skip connections.

In Fig. 4, extracted features or elements x_1^t to x_r^t are received by respective input nodes 410, conveyed to a layer 420 of LSTM blocks 422, from there to another layer 420 of the LSTM blocks 432 and then to a fully connected layer 440 and thereafter to an output layer

450 that outputs a predicted time until next break event of a given cable modem.

Additional network information

We can incorporate network information, such as geographical positions of the cable modems, amplifiers, and cable modems, in the algorithm in various ways; it can be, e.g.:

- 5 1. a separate predictor of the problems in the network, or
2. a joint predictor that uses both network information and temporal information from IPDR data.

The first approach is usable as follows, for example: (1) generate numerical embeddings of each vertex in the network graph, (2) pass these representations to the machine learning
10 as additional inputs, (3) use the machine learning model in the same way as the previously developed one.

In an example embodiment, a compute vertex embeddings are computed by taking corresponding rows from the adjacency matrix. For example, for graph represented in Fig.
5 the representation of a fifth node (denoted with reference sign 5) will be [0, 0, 1, 0, 0, 0,
15 0] (we interpret the illustrated graph as undirected).

The representation of the graph vertices can be replaced with any other approach that generates a meaningful representation of the graph vertices, such as node2vec or different modifications of biased random walks.

If the machine learning model works on top of aggregated data, the data is aggregated via
20 all the intervals with minimum, maximum, average, and percentile operations. Automatic feature extraction techniques can be performed in this step to improve predictive performance of the machine learning model.

Maintenance actions

The machine learning model based predictive maintenance system preferably works
25 continuously. It collects information from cable modem termination systems and other network elements. Newly collected data are pre-processed using the approach described in foregoing. With set intervals, such as once a day, the system makes predictions for all registered devices and marks devices with the top N, such as 20, closest predicted breaks as problematic. This threshold can also be chosen according to predictive performance.

30 In an example embodiment, the threshold decision may be made based on a cost function, depending on the price of action and potential loss that a particular problem could cause. For example, a threshold may be selected to give 80% precision (number of true positives / (number of true positives plus false positives)) of the predictions drops less than 80% 2

weeks reduce the threshold to match 80%, if the performance is better than 80% per week increase the number of predictions by 10%. Again, 80% and 10% were chosen for one embodiment., In other embodiments, there may be other costs and different adjustment of the parameters, for instance, by making every iteration of the algorithm cost-effective.

- 5 If a cable modem has been recently repaired, the system may place it in a quarantine or a state where that cable modem is not considered faulty unless issues with that piece of equipment persist for a time exceeding a threshold value. Otherwise, the system may automatically trigger a service ticket for the cable modem and mark that ticket by type "proactive." Such pieces of equipment can be taken to an initial analysis for verifying of the
- 10 problem. In the initial analysis, for example, following items may be checked: current and historical measurements related to the cable modem, amplifiers in the area, whether there were any tickets from the customer, other users for similar tickets, etc. If a problem is found, this prediction is marked as correct and that finding is propagated to the machine learning model.

15 Deployment of the system

The system can be deployed with cloud technologies, for example. Here we describe an example process of delivering the machine learning model from source code in a code repository to production. Initially, source code can be stored in a code repository for different services. When the system is deployed, the code is delivered to a Jenkins server (or other

20 systems with similar capabilities) and built into Docker containers, and then run as Airflow jobs. Predictions are stored in Splunk. The machine learning model or models retraining is carried out by API requests and stored in a storage, such as an S3-compatible storage. The newest machine learning model is retrieved by an inference server and used for repeated (e.g., daily) predictions. In another example embodiment, the system is used more

25 frequently than once a day or as a real-time tool.

The models can be separated by cable modem termination systems if those have significant differences in measurements and functioning. In an example embodiment, one model is used for all cable modem termination systems.

The difference between CMTSs measurements can be measured in at least two ways:

- 30 Case 1: data collected from all the cable modem termination systems suffices for training separate models: (i) train the models separately and a singular model, (ii) compare performance on hold-out dataset or by performing backtest validation, (iii) perform statistical tests, and (iv) compare results to determine a best suited model.

Case 1: data collected from all the cable modem termination systems does not suffice for training separate models: (i) transform collected data in a lower dimensional space (for example, using principal component analysis, PCA), (ii) perform multiple sample testing, for instance with maximum mean discrepancy (MMD), test, (iii) obtain p-values (for example, using permutations test on the kernel matrix extracted from MMD kernel matrix). This shows which cable modem termination systems are similar and allow preparing a single model for similar cable modem termination systems.

Fig. 6 shows a block diagram of an apparatus 600 according to an example embodiment. The apparatus 600 comprises a communication interface 610; a processor 620; a user interface 630; and a memory 640.

The communication interface 610 comprises in an embodiment a wired and / or wireless communication circuitry, such as Ethernet; Wireless LAN; Bluetooth; GSM; CDMA; WCDMA; LTE; and / or 5G circuitry. The communication interface can be integrated in the apparatus 600 or provided as a part of an adapter, card, or the like, that is attachable to the apparatus 600. The communication interface 610 may support one or more different communication technologies. The apparatus 600 may also or alternatively comprise more than one of the communication interfaces 610.

In this document, a processor may refer to a central processing unit (CPU); a microprocessor; a digital signal processor (DSP); a graphics processing unit; an application specific integrated circuit (ASIC); a field programmable gate array; a microcontroller; or a combination of such elements.

The user interface may comprise a circuitry for receiving input from a user of the apparatus 600, e.g., via a keyboard; graphical user interface shown on the display of the apparatus 600; speech recognition circuitry; or an accessory device; such as a headset; and for providing output to the user via, e.g., a graphical user interface or a loudspeaker.

The memory 640 comprises a work memory 642 and a persistent memory 644 configured to store computer program code 646 and data 648. The memory 640 may comprise any one or more of: a read-only memory (ROM); a programmable read-only memory (PROM); an erasable programmable read-only memory (EPROM); a random-access memory (RAM); a flash memory; a data disk; an optical storage; a magnetic storage; a smart card; a solid-state drive (SSD); or the like. The apparatus 600 may comprise a plurality of the memories 640. The memory 640 may be constructed as a part of the apparatus 600 or as an attachment to be inserted into a slot; port; or the like of the apparatus 600 by a user or by another person or by a robot. The memory 640 may serve the sole purpose of storing data

or be constructed as a part of an apparatus 600 serving other purposes, such as processing data.

A skilled person appreciates that in addition to the elements shown in Fig. 6, the apparatus 600 may comprise other elements, such as microphones; displays; as well as additional circuitry such as input/output (I/O) circuitry; memory chips; application-specific integrated circuits (ASIC); processing circuitry for specific purposes such as source coding/decoding circuitry; channel coding/decoding circuitry; ciphering/deciphering circuitry; and the like. Additionally, the apparatus 600 may comprise a disposable or rechargeable battery (not shown) for powering the apparatus 600 if external power supply is not available.

Fig. 7 shows a flow chart according to an example embodiment of a method for predictive cable modem termination system maintenance, comprising automatically:

710. continually collecting state data of a cable modem termination system;

720. constructing a machine learning dataset from the collected data, including:

730. accruing past observation data from a considering period from the collected state data;

740. predicting future observation data during a prediction period using a classification task will there be a problem with a particular piece of equipment during the prediction period and a regression task to predict time until a next problematic situation;

750. using the machine learning dataset by a machine learning model selected from a group of gradient boosting; extremely randomised trees; and neural networks; and including

760. incorporating spatial knowledge of the cable modem termination system;

770. using the machine learning with the constructed machine learning dataset to predict pieces of equipment prone to experience maintenance issues during the prediction period.

Any of the afore described methods, method steps, or combinations thereof, may be controlled or performed using hardware; software; firmware; or any combination thereof.

The software and / or hardware may be local; distributed; centralised; virtualised; or any combination thereof. Moreover, any form of computing, including computational intelligence, may be used for controlling or performing any of the afore described methods, method steps, or combinations thereof. Computational intelligence may refer to, for example, any of artificial intelligence; neural networks; fuzzy logics; machine learning; genetic algorithms; evolutionary computation; or any combination thereof.

Various embodiments have been presented. It should be appreciated that in this document, words comprise; include; and contain are each used as open-ended expressions with no intended exclusivity.

The foregoing description has provided by way of non-limiting examples of particular

implementations and embodiments a full and informative description of the best mode presently contemplated by the inventors for carrying out the invention. It is however clear to a person skilled in the art that the invention is not restricted to details of the embodiments presented in the foregoing, but that it can be implemented in other embodiments using
5 equivalent means or in different combinations of embodiments without deviating from the characteristics of the invention.

Furthermore, some of the features of the afore-disclosed example embodiments may be used to advantage without the corresponding use of other features. As such, the foregoing description shall be considered as merely illustrative of the principles of the present
10 invention, and not in limitation thereof. Hence, the scope of the invention is only restricted by the appended patent claims.

CLAIMS

1. A method for predictive cable modem termination system maintenance, comprising automatically:

continually collecting (710) state data of a cable modem termination system;

5 constructing (720) a machine learning dataset from the collected data, including:

accruing (730) past observation data from an information period from the collected state data;

predicting (740) future observation data during a prediction period using a classification task will there be a problem with a particular piece of equipment during the prediction period and a regression task to predict time until a next problematic situation;

10 using (750) the machine learning dataset by a machine learning model selected from a group of gradient boosting; extremely randomised trees; and neural networks; and

incorporating (760) spatial knowledge of the cable modem termination system;

the method further comprising using (770) the machine learning with the constructed machine learning dataset to predict pieces of equipment prone to experience maintenance issues during the prediction period;

15 wherein the predicting of the pieces of equipment prone to experience maintenance issues during the prediction period comprises choosing the machine learning model from two or more alternatives using a cross-validation.

20 2. The method of claim 1, wherein the predicting of the pieces of equipment prone to experience maintenance issues during the prediction period comprises determining a set of most likely pieces of equipment to experience the maintenance issues.

3. The method of claim 1 or 2, wherein the predicting of the pieces of equipment prone to experience maintenance issues during the prediction period comprises determining a precision of the predicting by a ratio of a number of correct positives divided by a number of all predicted pieces of equipment.

4. The method of claim 3, wherein the predicting of the pieces of equipment prone to experience maintenance issues during the prediction period comprises adapting the given number of most likely pieces of equipment according to development of the ratio.

30 5. The method of any one of preceding claims, wherein the state data further comprises service tickets issued by user as an indicator of breaks or other issues in the cable modem termination network.

6. The method of any one of preceding claims, wherein the state data comprises one or

more of following: a signal to noise ratio; a number of corrected codewords; a cable upstream equalization-coefficient; an upstream high resolution timing offset which represents a round trip time on an upstream channel of a cable modem; a Boolean flag for muting state of the channel; a number of microreflections received on an interface; a ranging status of a cable modem; a receive power as perceived for the upstream channel; a number of codewords received with uncorrectable errors from the cable modem; a number of codewords received without any errors; and / or a system uptime value taken from the cable modem termination system when the Internet Protocol Detailed Record, IPDR, was created.

7. The method of any one of preceding claims, wherein a plurality of upstream channels are jointly represented.

8. The method of any one of preceding claims, wherein the training of the machine learning model is performed using a binary classification task and a regression task.

9. The method of any one of preceding claims, wherein a joint predictor is computed using both network information and temporal information from the collected state data.

10. The method of any one of preceding claims, wherein verification of predicted issues is propagated back to the machine learning model.

11. The method of any one of preceding claims, wherein one machine learning model is used for two or more cable modem termination systems.

12. The method of any one of preceding claims, wherein pieces of equipment comprise one or more amplifiers and one or more cable modems.

13. A computer program (646) comprising computer executable program code which when executed by at least one processor (620) causes an apparatus (600) at least to perform the method of any one of preceding claims.

14. An apparatus (600) comprising means (610, 620, 630, 640) for performing the method of the any one of claims 1 to 12.

PATENTTIVAATIMUKSET

1. Menetelmä kaapelimodeemiterminointijärjestelmän ennakoiwaan ylläpitoon, jossa automaattisesti:

jatkuvasti kerätään (710) kaapelimodeemiterminointijärjestelmän tilatietoja;

5 rakennetaan (720) koneoppimisaineisto kerätyistä tiedoista, sisältäen sen, että:

kerätään (730) aiempia havaintotietoja tietojaksolta kerätyistä tilatiedoista;

ennustetaan (740) tulevia havaintotietoja ennustejakson aikana käyttäen luokitustehtävää, että onko tietyssä laitteessa ongelma ennustejakson aikana ja regressiotehtävää ennustamaan ajan seuraavaan ongelmalliseen tilanteeseen;

10 käytetään (750) koneoppimistietojoukkoa koneoppimismallilla, joka on valittu joukosta, jonka muodostavat gradienttitehostus, äärisatunnaistetut puut; ja hermoverkot;

sisällytetään (760) kaapelimodeemiterminointijärjestelmän paikkatietoa;

johon menetelmä lisäksi kuuluu (770) koneoppimisen käyttö rakennetun koneoppimisen tietojoukon kanssa sellaisten laitteiden ennustamiseksi, jotka ovat alttiita ylläpito-ongelmille ennustejakson aikana;

15 jossa ennustejakson aikana ylläpito-ongelmille alttiiden laitteiden ennustaminen käsittää koneoppimismallin valitsemisen kahdesta tai useammasta vaihtoehdosta käyttäen ristiinvalidointia.

2. Patenttivaatimuksen 1 menetelmä, jossa ylläpito-ongelmille ennustejakson aikana alttiiden laitteiden ennustaminen käsittää todennäköisimmin ylläpito-ongelmia kokevien laitteiden joukon määrittämisen.

3. Patenttivaatimuksen 1 tai 2 mukainen menetelmä, jossa ylläpito-ongelmille ennustejakson aikana alttiiden laitteiden ennustaminen käsittää ennustamisen tarkkuuden määrittämisen suhteella oikeiden positiivisten määrä jaettuna kaikkien ennustettujen laitteiden määrällä.

4. Patenttivaatimuksen 3 menetelmä, jossa ylläpito-ongelmille ennustejakson aikana alttiiden laitteiden ennustaminen käsittää tietyn todennäköisimpien laitteiden määrän mukauttamisen suhteen kehityksen mukaan.

5. Minkä tahansa edeltävän patenttivaatimuksen menetelmä, jossa tilatiedot sisältävät edelleen käyttäjän lähettämät palveluliput indikaattorina katkoista tai muista ongelmista kaapelimodeemiterminointiverkossa.

6. Minkä tahansa edeltävän patenttivaatimuksen menetelmä, jossa tilatiedot sisältävät yhden tai useamman seuraavista: signaali-kohinasuhde; korjattujen koodisanojen määrä;

kaapelin ylävirran ekvalisointivakio; ylävirran korkean resoluution ajoituspoikkeama, joka edustaa edestakaista aikaa kaapelimodeemin ylävirran kanavalla; Boolean-lippu kanavan mykistystilalle; rajapinnalla vastaanotettujen mikroreflektioiden määrä; kaapelimodeemin ranging-tila; ylävirran kanavalla havaittu vastaanottoteho; kaapelimodeemista saatujen korjaamattomissa olevien virheiden määrä; koodisanojen määrä, jotka on vastaanotettu ilman virheitä; ja / tai järjestelmän käytettävyyssika, joka on otettu kaapelimodeemiterminointijärjestelmästä, kun Internet Protocol Detailed Records, IPDR, -tietue luotiin.

7. Minkä tahansa edeltävän patenttivaatimuksen menetelmä, jossa useita alkupään kanavia on yhdessä edustettuna.

8. Minkä tahansa edeltävän patenttivaatimuksen menetelmä, jossa koneoppimismallin koulutus suoritetaan binäärisen luokittelutehtävän ja regressiotehtävän avulla.

9. Minkä tahansa edeltävän patenttivaatimuksen menetelmä, jossa yhteinen ennustaja lasketaan käyttäen sekä verkkotietoja että hetkellistä tietoa kerätyistä tilatiedoista.

10. Minkä tahansa edeltävän patenttivaatimuksen menetelmä, jossa ennustettujen ongelmien todentaminen propagoidaan takaisin koneoppimismalliin.

11. Minkä tahansa edeltävän patenttivaatimuksen menetelmä, jossa yhtä koneoppimismallia käytetään kahdessa tai useammassa kaapelimodeemiterminointijärjestelmässä.

12. Minkä tahansa edeltävän patenttivaatimuksen menetelmä, jossa laitteet käsittävät yhden tai useamman vahvistimen ja yhden tai useamman kaapelimodeemin.

13. Tietokoneohjelma (646), joka käsittää tietokoneella suoritettavaa ohjelmakoodia, joka vähintään yhden prosessorin (620) suorittamana aikaansaa laitteen (600) suorittamaan ainakin minkä tahansa edeltävän patenttivaatimuksen menetelmän.

14. Laite (600), joka käsittää välineet (610, 620, 630, 640) minkä tahansa patenttivaatimuksen 1–12 menetelmän suorittamiseksi.

Fig. 1

The diagram illustrates a system 100 for processing data from multiple cameras and a user. The system includes the following components and connections:

- CM 1** (110) and **CM 2** (110') are cameras that provide input to the **Amp** (114) and the **Data collector** (130).
- USR** (116) is a user interface component that provides input to the **Ticketing system** (150).
- The **Amp** (114) receives input from CM 1 (110) and CM 2 (110') and outputs data to the **Data collector** (130) via connection 122.
- The **Data collector** (130) receives input from CM 1 (110) via connection 112 and outputs data to the **Database** (140) via connection 132.
- The **Database** (140) is a central storage component that provides data to the **Ticketing system** (150) via connection 152, the **Predictor** (160) via connection 142, and the **ML model training** (170) via connection 144.
- The **Predictor** (160) provides input to the **Ticketing system** (150) via connection 162.
- The **ML model training** (170) provides input to the **Database** (140) via connection 172.

The diagram illustrates a network architecture. A central cloud labeled "INTERNET" is connected to three external components: a "Controller" (labeled 210), "Kafka" (labeled 220), and "Storage" (labeled 230). Each component is represented by a rectangular box. The connections are shown as lines extending from the boxes to the cloud boundary. A reference numeral 240 points to the cloud boundary.

Fig. 3

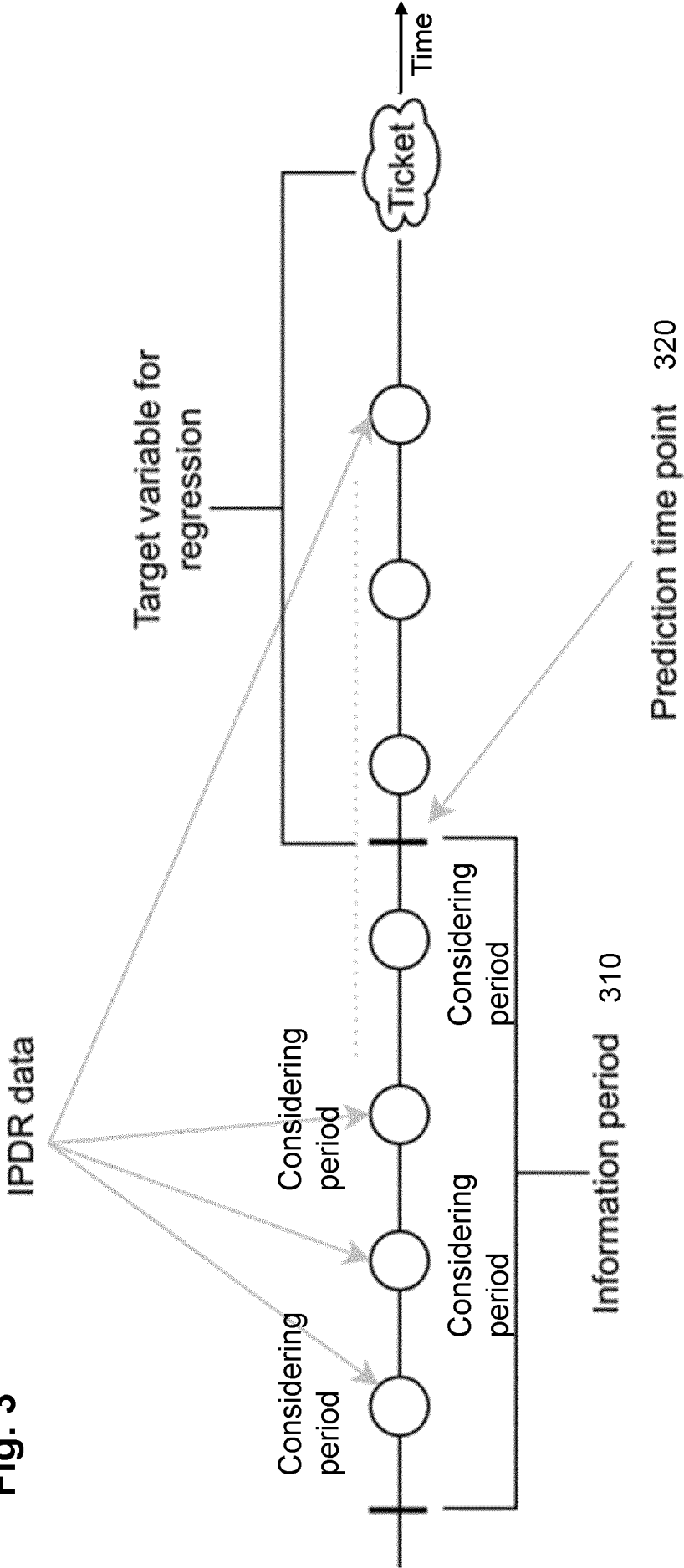


Fig. 4

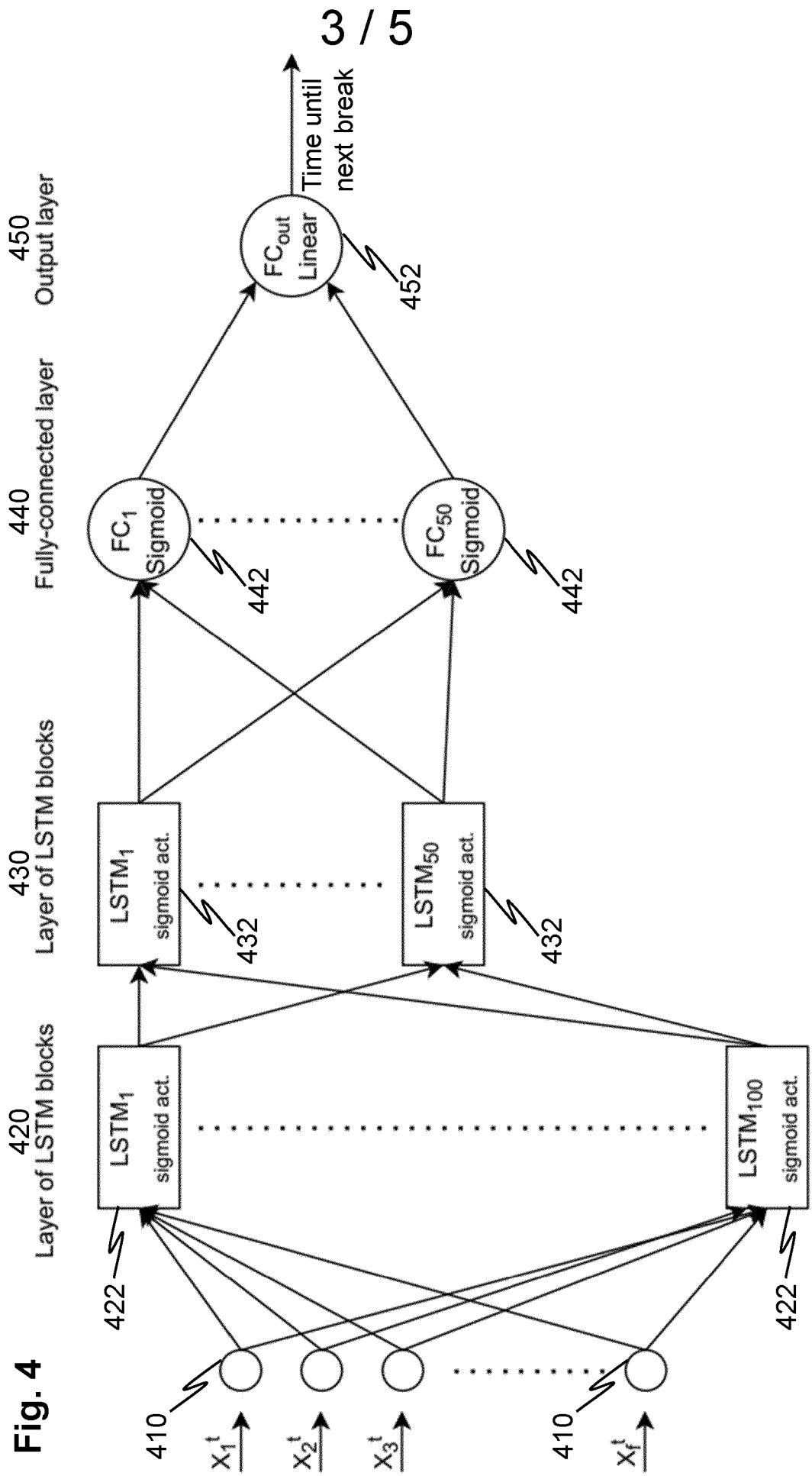


Fig. 5

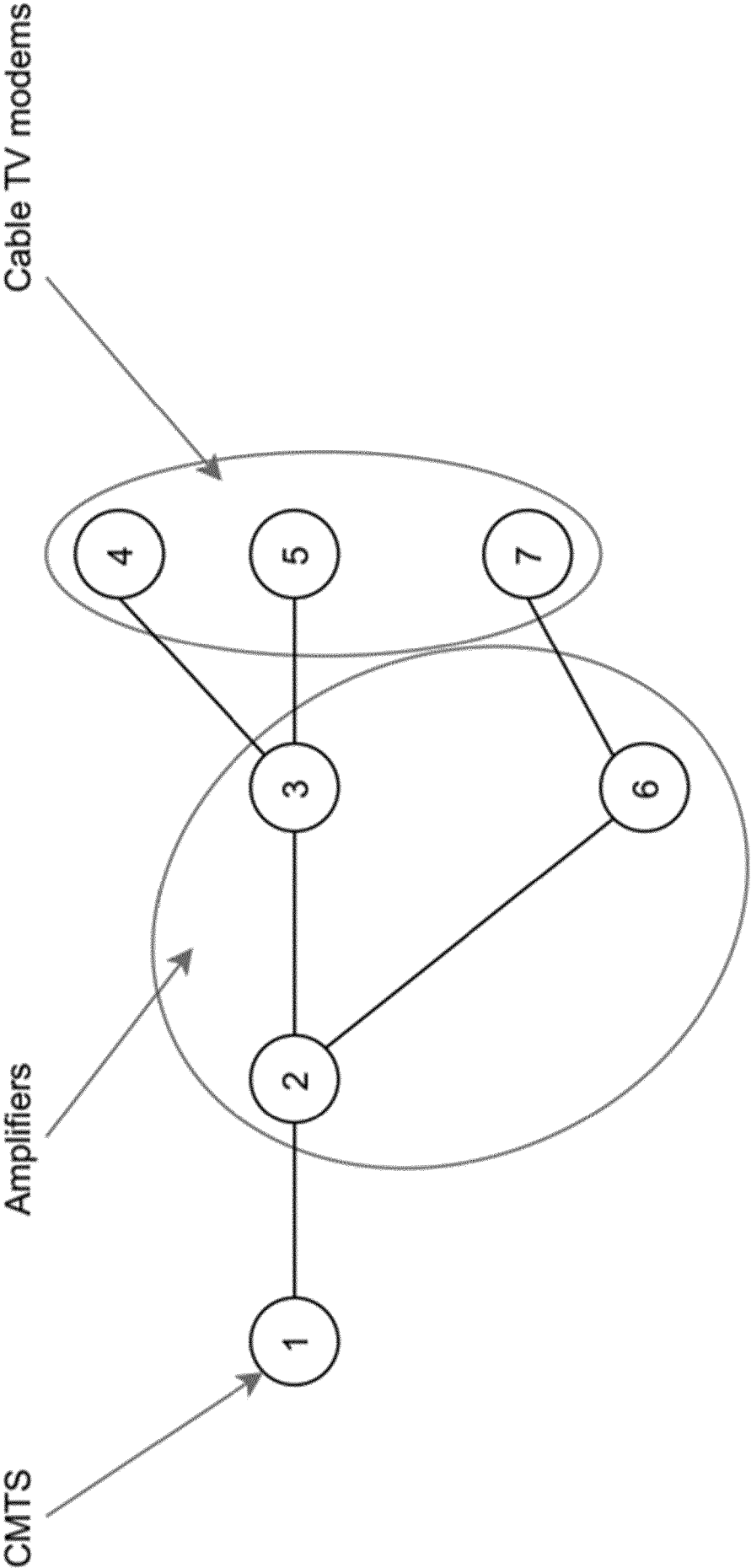
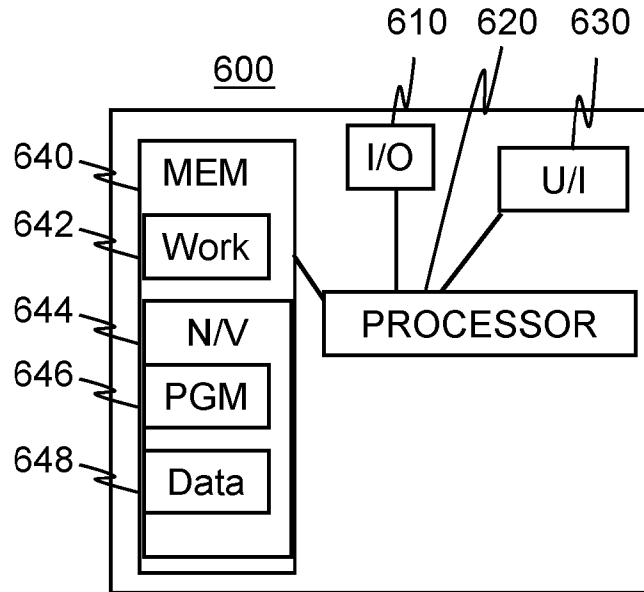


Fig. 6**Fig. 7**

710. continually collecting state data of a cable modem termination system

720. constructing a machine learning dataset from the collected data, including:

730. accruing past observation data from an information period from the collected state data

740. predicting future observation data during a prediction period using a classification task will there be a problem with a particular piece of equipment during the prediction period and a regression task to predict time until a next problematic situation

750. using the machine learning dataset by a machine learning model selected from a group of gradient boosting; extremely randomised trees; and neural networks; and including

760. incorporating spatial knowledge of the cable modem termination system

770. using the machine learning with the constructed machine learning dataset to predict pieces of equipment prone to experience maintenance issues during the prediction period