ANALÝZA ALARMOV CESTNÉHO TUNELA

ROAD TUNNEL ALARM LOG ANALYSIS

Igor Miklóšik[1], Tomáš Tichý[2], Jiří Štefan[3]

ABSTRAKT

Článok sa zaoberá analýzou alarmov tunela použitím zhlukovej analýzy, analýzy časových radov a metódami detekcie anomálií. Tieto algoritmy by mali varovať technikov údržby pred zvýšeným výskytom porúch konkrétneho technologického komponentu alebo skupiny komponentov. Do úvahy je tiež vzatá hierarchia zapojenia komponentov pre detekciu a elimináciu porúch spôsobených komponentami na vyššej úrovni hierarchie.

ABSTRACT

The paper concerns an analysis of tunnel alarm logs using clustering, time series prediction and anomaly detection methods. These algorithms should warn the tunnel maintenance before an increased occurrence of malfunctions of a particular technological component or group of components. Hierarchy of the components is also considered to detect and eliminate malfunctions caused by components at higher hierarchy level.

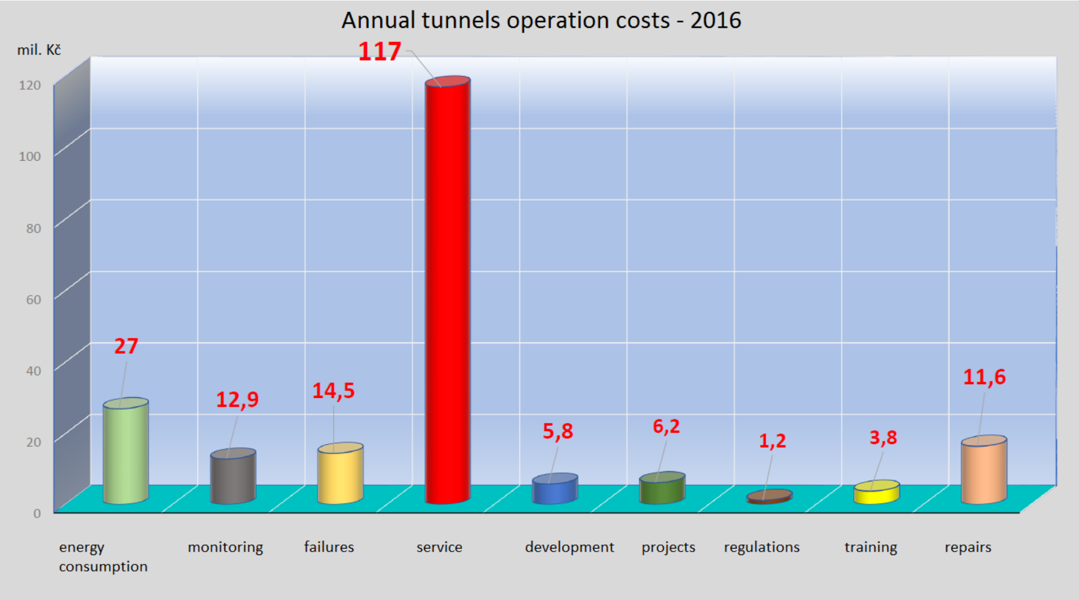
1 Introduction

Road tunnel is operated in real-time therefore equipment lifetime should be seriously monitored to provide expected safety level. Lifetime of equipment varies according to PIARC [1] from 10 to 25 years. The results demonstrated big differences: lowest average lifetime (10 years) was observed at electronic systems, Illumination and SCADA monitoring equipment, longest average lifetime was observed at energy supply and cabling (from 20 – 25 years). Ventilation, Signing and Safety equipment lifetime was between 15 and 20 years, deviations for all mentioned subsystems was between 3 and 5 years. Costs for service are according to experience in Czech Republic highest part of the tunnel operational costs. Situation from the year 2016 is illustrated on Obr. 1. It proves that installed technological components have to be unified, service and maintenance have to be standardized and optimized together with failures prediction. According to [3] tunnel operating and maintenance paper for Czech Republic, technical condition of the tunnel construction and equipment is divided into 7 discrete steps and should be considered during regular revisions. Recommended periodicity of these revisions is also specified in the same document. Predictions of the tunnel construction conditions can be performed according to ISO 15686-2 [4] as polynomial extrapolation of values obtained from experts.

[1] Ing. Igor Miklóšik, PhD., ELTODO SK, a.s., M.R. Štefánika 73, 010 01 Žilina, e-mail: miklosikj@eltodo.sk

[2] doc. Ing. Tomáš Tichý, PhD. MBA., ELTODO, a.s., Novodvorská 1010/14, Praha 4, e-mail: tichyt@eltodo.cz

[3] Ing. Jiří Štefan, ELTODO, a.s., Novodvorská 1010/14, 142 00 Praha 4, e-mail: stefanj@eltodo.cz

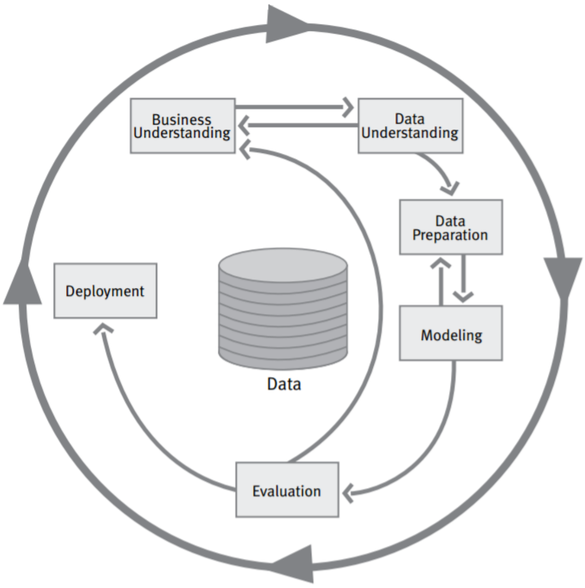


Obr. 1 Ročné prevádzkové náklady tunelov [2]
Fig. 2 Annual tunnels operation costs [2]

Since tunnel technological equipment consists of heterogeneous subsystems [5] and they could be replaced independently, different approach to construction conditions prediction should be used. We decided to use alarm log analysis.

2 Data Analysis

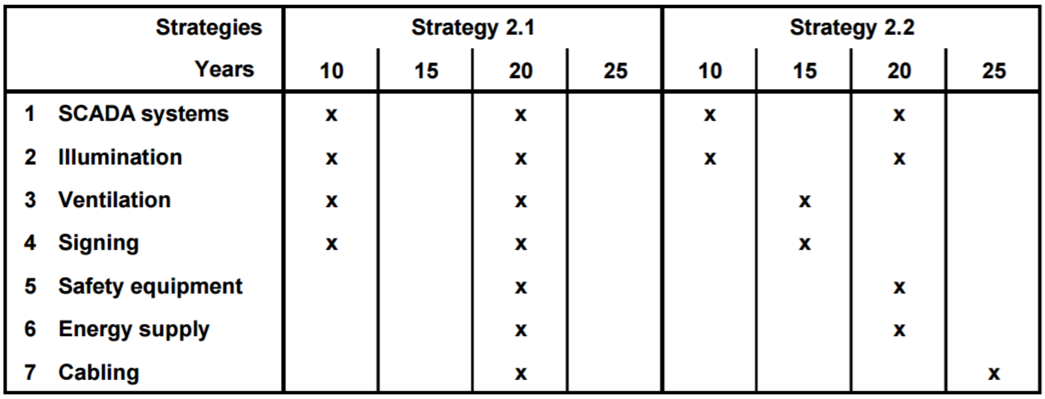
Data analysis has been performed under CRISP-DM principle (Cross Industry Standard Process for Data Mining) [6]. The reason for selection of this principle was that our goal was also focus on business aspects, not only on modeling aspects. Therefore SEMMA principle (Sample, Explore, Modify, Model, Assess) has not been used. CRISP-DM principle consists of six steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. Whole process is illustrated on Obr. 3 ***Kroky CRISP-DM [6]***
Fig. 4.



Obr. 3 Kroky CRISP-DM [6]
Fig. 4 Steps of CRISP-DM [6]

3 Business Understanding

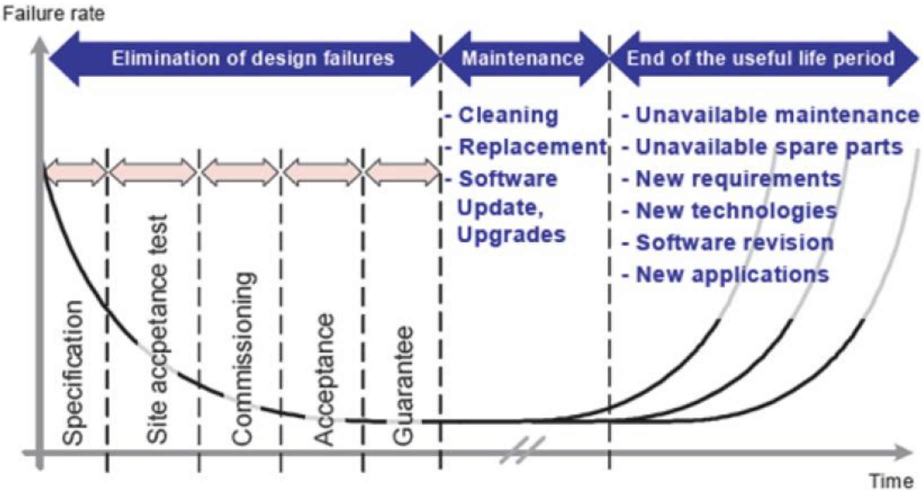
The most common requirement is to provide the statistics from the tunnel operation, which can be realized by known tools such as Microsoft EXCEL or Microsoft Power BI [7]. Our business idea was to extend the statistics with additional information to optimize service costs, number of components in the stock, warnings before problematical components series or producers and help the operators to handle the alarm situations.



Strategies	Strategy 2.1				Strategy 2.2			
Years	10	15	20	25	10	15	20	25
1 SCADA systems	x		x		x		x	
2 Illumination	x		x		x		x	
3 Ventilation	x		x			x		
4 Signing	x		x			x		
5 Safety equipment			x				x	
6 Energy supply			x				x	
7 Cabling			x					x

Obr. 5 Stratégie údržby [1]
Fig. 6 Maintenance strategies [1]

Based on the observed average lifetimes mentioned before two maintenance strategies can be used: replacement of the components immediately after the failure or replacement according to the plan. Comparison of replacement costs in [1] demonstrated that planned maintenance is economically more efficient. Two strategies with planned replacements are on Obr. 5. Useful lifetime of the components can be seen on the well-know bathtube curve from reliability engineering displayed on Obr. 7. Our attention is concentrated on the "Maintenance" part and "End of the useful life period" part, especially on the change-over between them. This is indicated by the higher failure rates of the components, which directly leads to higher service costs.



Obr. 7 Vaňová krivka [1]
Fig. 8 Bathtube curve [1]

4 Data Understanding

Example records from Ventilation and Illumination tunnel alarm logs are in Tab. 2.

Tab. 1 Príklad alarmov tunela
Tab. 2 Tunnel alarm data example

Field	Ventilation	Illumination
Time	1.10.2010 9:51:59	1.1.2011 0:05:42
Sequence number	12736620	15486482
Alarm ID	V.VR.VRV01.ERR02	O.AO.AO051.COM_FAIL1
Alarm class	A2	A2
Group	VENTILATION	ILLUMINATION
Logged by	MASTER_PTM0_RP	MASTER_PTM0_RP
Reference	V.VR.VRV01.ERR02	O.AO.AO051.COM_FAIL1
Previous state	G	G
Logged state	R	R
Final state	R	R
Alarm message	IV0-V-VE00001: Command execution error	IL0-O-LA00051: XCOM1 communication error
Time of generation	1.10.2010 9:51:52	1.1.2011 0:05:38

Legend of technological components marking is described in [8] XYY-Z-ABCCCCC:

- X - tunnel name,
- YY - location, tunnel tube, traffic lane,
- Z - technological group,
- AB - device marking,
- CCCCC - location, numbering of the component.

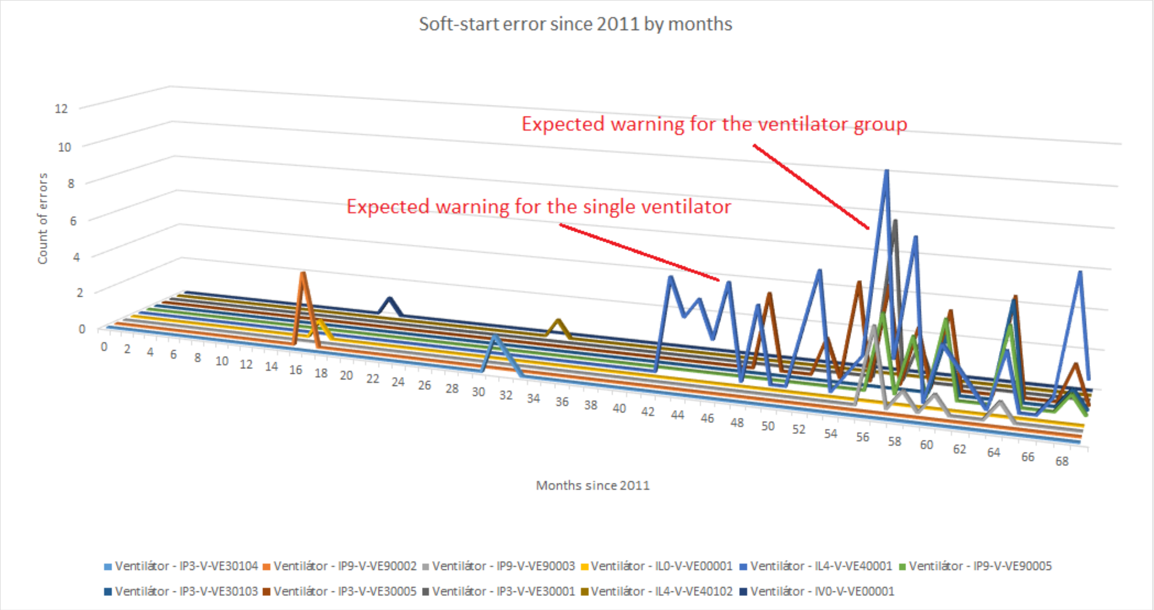
In our case for the illumination example IL0-O-LA00051 from the Tab. 2 (*Alarm message* field) is marking as follows:

- I - tunnel Cholupice,
- L - left tunnel tube, 0 - internal traffic lane,
- O - Illumination,
- LA - adaptation (accommodation) lighting,
- 00051 - Numbering.

The main problem discovered during the data analysis was that maintenance process did not interconnect directly the service action (repair, configure, replacement) with the specific component. This information is stored in many pdf documents, manual viewing and assigning service actions to specific components to obtain time between failures is a time-consuming task. There are few components with operating hours stored in the SCADA system such as ventilators or lighting and operator should reset them manually after their replacement. Even though this procedure is recommended, it's not performed for sure in every case. Also the values stored in the PLC are not always recovered after the reboot in every installation, so reliability analysis could not be easily performed as a well-known and accepted method. At the current state of the data structure we decided to use Time series analysis as the optimal approach. For the future, a kind of ticketing system is recommended to simplify and clarify the maintenance process. Selection of the service action performed on each component would be forced by the proper field in electronic form and service history would be traceable for each installed component. Then reliability analysis could be performed and experience with the components in each tunnel could be stored in global database for the maintenance and stock optimization.

5 Data Preparation

We were interested only in situation when alarms were reset because we wanted to count their time durations. So we filtered only alarms with the fields *Previous state* set to G (Generated) and *Final state* set to R (Reset). Then we subtracted the *Time field* and *Time of generation* and we obtained the alarm duration. We omitted following fields from the Tab. 2: *Sequence number, Alarm ID, Logged by, Reference, Previous state, Logged state* and *Final state*. We wanted to assign the alarms to specific component, the component name is separated by a colon in the *Alarm message* field. Technicians usually generate undesirable alarms during service, installation and testing the components functionality. So we filtered alarms generated during the maintenance time of the tunnel. We filtered out the alarms from the other technological groups except Ventilation and Video-detection, due to robustness of the data. Significant alarms growth in both mentioned technological groups required additional maintenance intervention in the past. Problems with soft-starters of the ventilators occurred in ventilation group and communication failures with cameras occurred in video-detection group. We decided to analyze the soft-starters case in detail in following chapters to prevent the additional intervention. The situation with soft-starters alarms history with expected warnings for the maintenance is on Obr. 9.

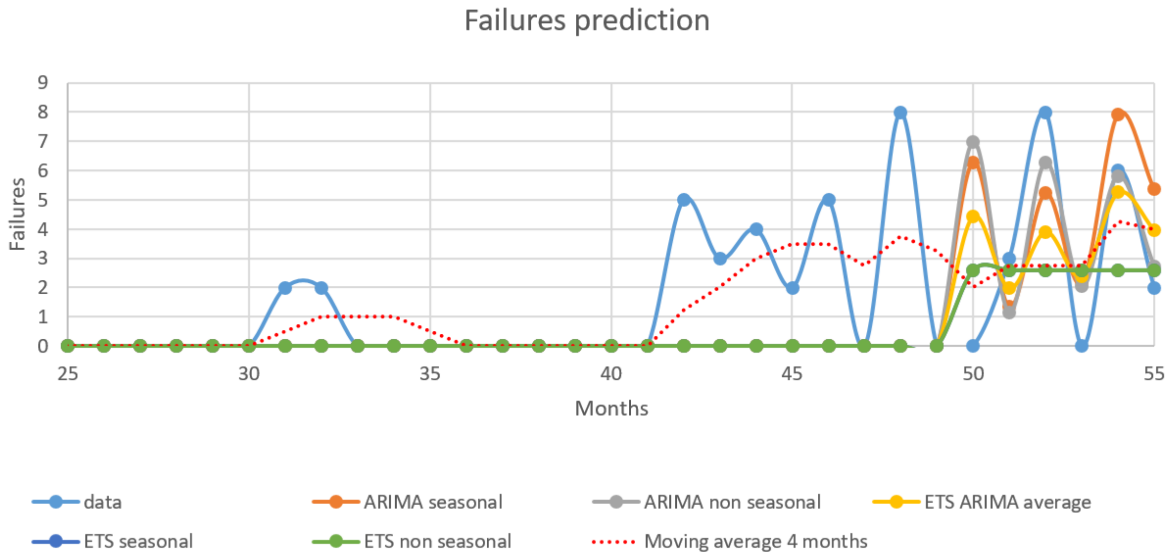


Obr. 9 Poruchy Soft-štart obvodov
Fig. 10 Soft-start circuit errors

6 Modelling – Prediction

As mentioned before time series analysis has been selected to predict the value of alarms of technological components. Following models have been considered: ARIMA (Autoregressive integrated moving average) and ETS (Exponential smoothing). Both seasonal and non-seasonal variants of the models have been evaluated although alarms generated during regular tunnel maintenance should be filtered. Traffic intensity has also seasonal trends during the year and can influence the results. Seasonal variant models should therefore use seasonality value set to 12 months. Non-seasonal ARIMA(p,d,q) models use three parameters p, d, and q (non-negative integers), p is the order of the autoregressive model, d is the degree of differencing, and q is the order of the moving-average model. Seasonal ARIMA(p,d,q)(P,D,Q)$_m$, use in addition to non-seasonal model also seasonal part marked with the uppercase letters P,D,Q, and m refers to the number of periods in each season.

We performed the prediction in R language from the 49. month (since 2011) on Obr. 11 to prevent the situation with rising failures in later months.



Obr. 11 Predikcia počtu porúch
Fig. 12 Failures prediction

We compared predicted values with real failures using Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Scaled Error (MASE) and Symmetric Mean Absolute Percentage Error (sMAPE). ARIMA models predicted better oscillating behavior between months than ETS models, seasonal ARIMA model also better predicted increasing trend of the failures.

7 Modelling – Anomaly detection

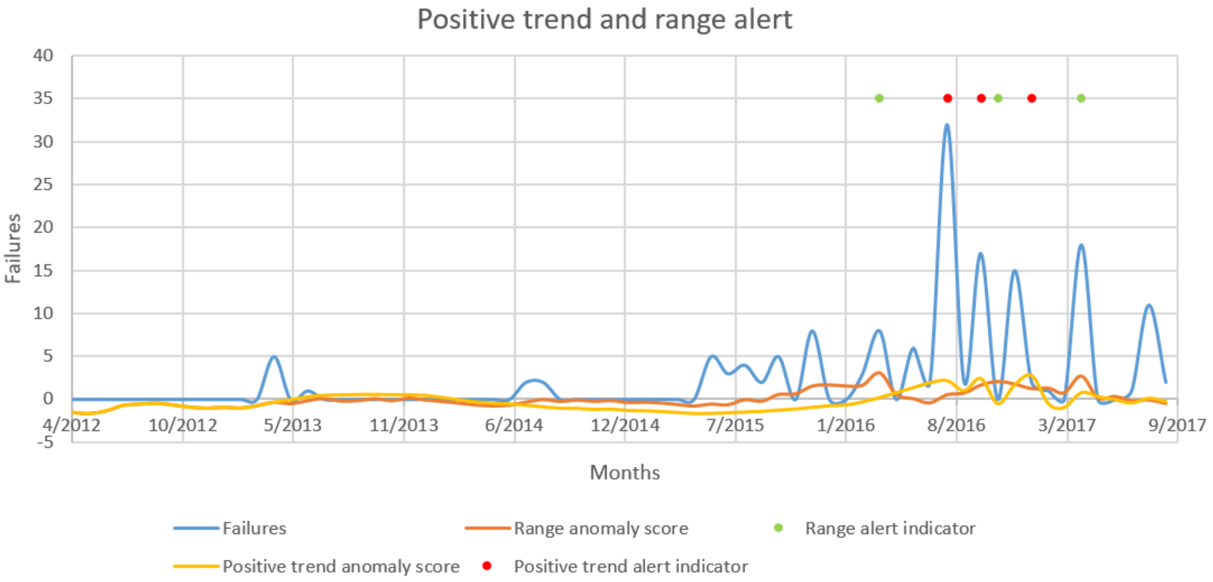
Anomaly detection in time series helps to detect abnormal behavior in real-time processes, in our case count of failures of technological components. Several algorithms can be used to achieve this goal such as: Twitter Seasonal Hybrid ESD (S-H-ESD) [9], Azure anomaly detection [10], Seasonal-trend decomposition, ARIMA, ETS and others. We have selected Time series anomaly detection algorithm in Azure for our tests with tunnel alarm logs. Algorithm can detect different types of anomalies on time series data:

- *Positive and negative trends*: we are interested in positive trend of the failures,
- *Increase in the dynamic range of values*: we are also interested in level changes.

Anomaly detection algorithm parameters:

- *Martingale Type*: Martingale function PowerAvg or Power,
- *Strangeness Function Type*: RangePercentile, SlowPosTrend and SlowNegTrend,
- *Length of Martingale and Strangeness Values*,
- *Alert Threshold*.

Martingale function PowerAvg without any additional parameters should have stable detector suitable for most needs. Power martingale function provides the Epsilon parameter from 0 to 1 to specify the sensitivity of the detector. *Strangeness Function Type* has been tested with RangePercentile and SlowPosTrend options to detect the growth of the failures. *Length of Martingale and Strangeness Values* parameters can specify the length of history window used to compute the strangeness (data points needed to learn "normal" behavior). We configured them to the same value as recommended. *Alert Threshold* value is used to generate the alert if counted anomaly score is above specified threshold. We can see from the Obr. 13 that selected thresholds should be lowered, since positive trend alert (red dots) is generated too late to prevent occurrence of high count of failures. Range anomaly score increases earlier during occurrence of the first repetitive failures therefore range alert (green dots) is generated earlier. So combination of different alert types is desirable to reach reliable anomaly detection.

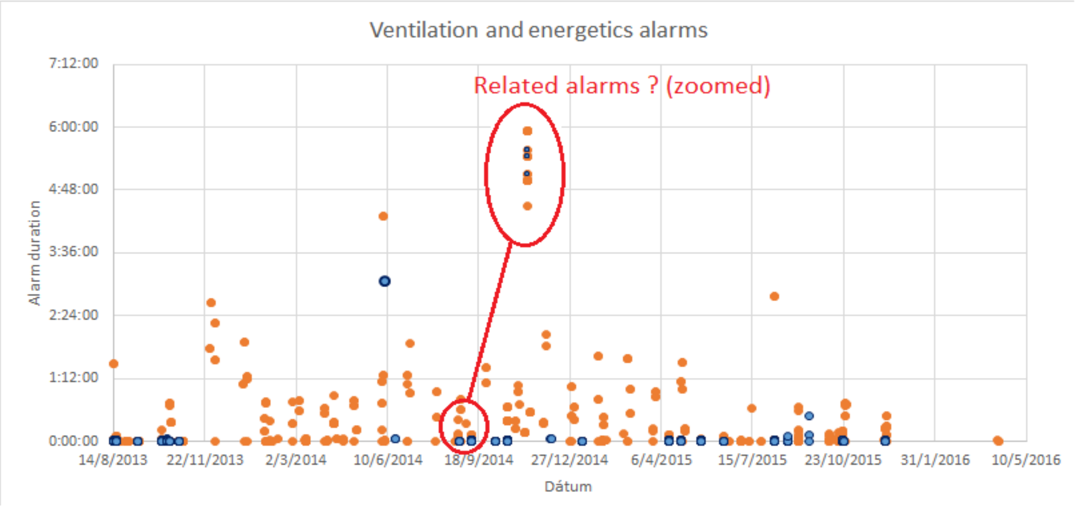


Obr. 13 Detekcia anomálií
Fig. 14 Anomaly detection

8 Modelling – Clustering

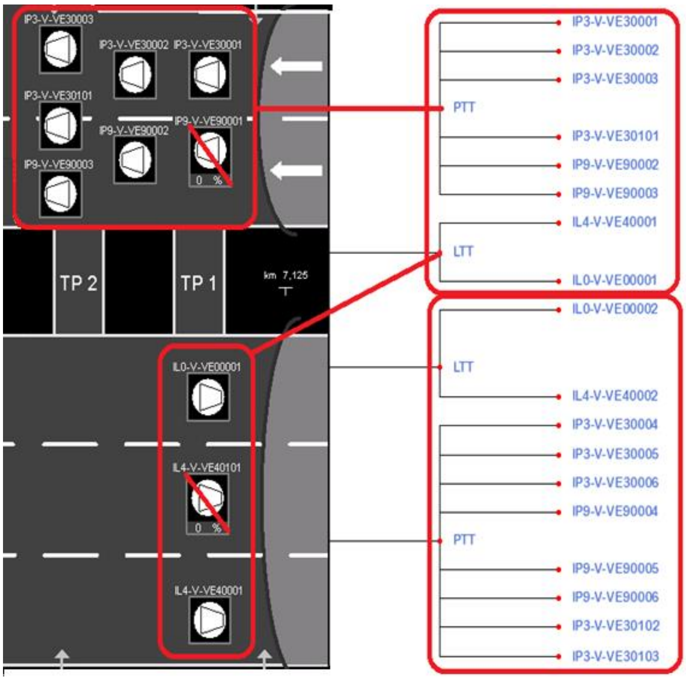
Cluster is in our case understood as a group of components which generate alarms in the logs at short time interval. Time interval for one cluster should be set to several seconds (0 - 20) due to hardware delays and hierarchical interconnection of the components. This situation is especially visible during the power failures, alarms are generated by all components connected to the non-functioning power supply components. Generation of many alarms may confuse the operator to localize the main cause of the failure. Therefore it is appropriate to identify the component or alarm with the highest hierarchy in the cluster.

Popular K-means algorithm from centroid clustering models cannot be used, it represents each cluster by a single mean vector. We cannot specify in advance how many clusters are present in the alarms. Also connectivity models and hierarchical clustering, which analyze the distance between items in cluster, have been considered. The tree representation was in our case not readable enough due to extensive data file and the selection of the tree level for optimal clusters representation was not straightforward. Our preferable solution specified the minimal density of the alarms included into the cluster and allowed omitting orphans (items not included in the clusters). This could be performed in density models e.g. DBSCAN algorithm. Time occurrence and duration of the alarms from ventilation group and energetics is on Obr. 15.



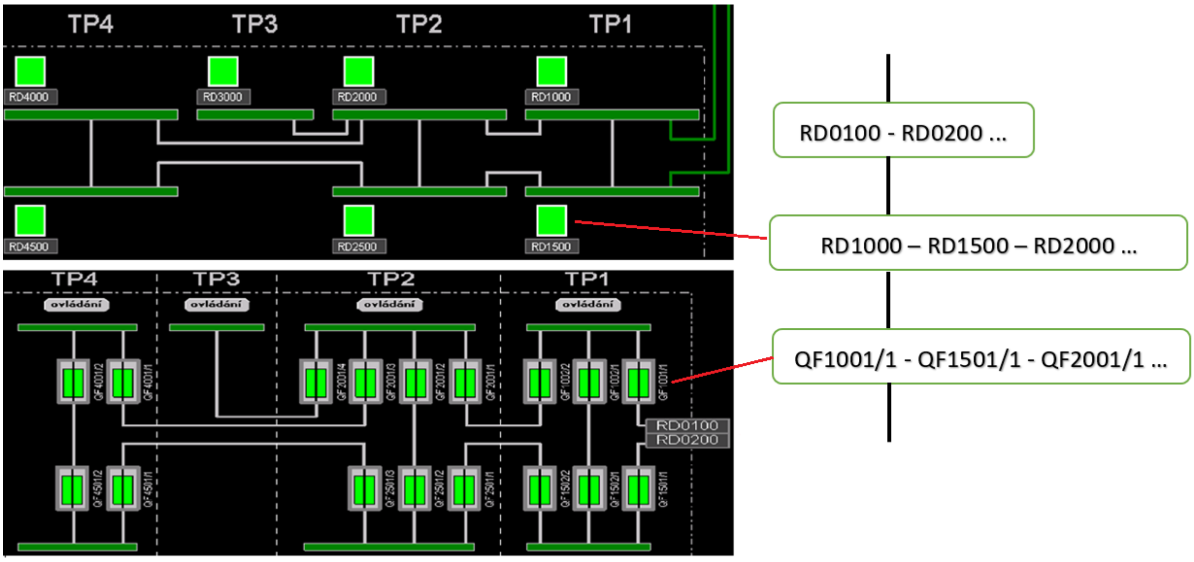
Obr. 15 Alarmy vetrania a energetiky
Fig. 16 Ventilation and energetics alarms

DBSCAN algorithm expects two input parameters: ε and minimum density. In our case ε is represented by the time interval and is expected to be within 5 - 20 seconds range. Minimum density represents the count of alarms in one cluster and in our case is equal to three or greater. We have tested also the value two for minimal alarms count, but DBSCAN generated many unwanted clusters with only one device. The reason for this is that many devices have special separate cumulative alarm generated together with the source alarm for the same device. Since failures occur many times duplicate clusters were recognized and frequency analysis had to be performed to obtain only relevant clusters.



Obr. 17 Clustre vetranie
Fig. 18 Clusters ventilation

Clusters containing ventilators for both left and right tunnel tube are on ***Obr. 17***. It can be seen that ventilators are inserted into the clusters correctly according to their location in the tunnel. Two ventilators crossed with red were equipped with different hardware (frequency converters) and alarms occurred few seconds later. We are interested in alarms when at least one technological group is energetics; these show us interconnections between components.



Obr. 19 Clustre energetika – redundancia zapojenia
Fig. 20 Clusters energetics – connection redundancy

The clusters containing power supply components are on Obr. 19. It can be seen on the left part of the figure that redundancy of the interconnections between the components provide higher tolerance to power failures. That is desirable and reasonable feature. On the other hand, this includes many components into one large cluster on the same hierarchy level. So we cannot generate the connection tree with a node for each component, we can generate only the tree with node for group of components on the same hierarchy level.

9 Deployment

Tunnel is specific environment because provider can require its operation in a private network for safety reasons. On-line cloud storage of the alarms such as Microsoft Azure and Machine Learning Studio [12] should not be used. Although on-line cloud storage at least of the models is advantageous to allow easier maintenance and adjustment as we tested with mentioned products. We decided to use the R language [11] because it provides many libraries and can be easily integrated into both online and offline common BI tools. Modern tools solved problems with robustness of the data in EXCEL in the past, they can provide comfort interactive interface to display the alarm history of proper components. Integration of the R language in the tools allow usage of many advanced algorithms.

10 Conclusion

We have proposed new approach for obtaining the additional information from the tunnel alarm logs. It is inspired by the idea of helping the operators and technicians to optimize the maintenance and reduce costs. Previous works in the field of tunnel data analysis was limited to the common statistical indicators. Although we did not obtain all information we expected (precise hierarchy of all components), our analysis showed that our approach was able to improve the information value of the alarm reports, warn the operators before the problematical components and indicate the necessity of earlier regular maintenance. Since the project is still under development, cost analysis has not been performed to demonstrate the costs reduction for each maintenance strategy. In the future we intend to use also reliability analysis after obtaining the complete service data for each component.

Bibliography

1. PIARC, Life cycle Aspects of Electrical Road Tunnel Equipment. PIARC Technical Commitee C.4, 2012.
2. F. Rainer, Záměr pro unifikaci stavební a technologické části tunelů na pozemních komunikacích ve správě ŘSD ČR. Konference Požární bezpečnost tunelů, Rožňov pod Radhošťem, 2017.
3. ELTODO, TP154 Provoz, správa a údržba tunelů pozemních komunikací Ministerstvo dopravy ČR, 2001.
4. ISO, ISO 15686 - Building and constructed assets Service life planning, Part 2: Service life prediction procedures. Technical Committee ISO/TC 59, 2009.
5. ELTODO, TP98 Technologické vybavení tunelů pozemních komunikací. Ministerstvo dopravy ČR, 2004.
6. Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer, and Wirth, CRISP-DM 1.0, Step-by-step data mining guide. SPSS, 2000.
7. Microsoft, Power BI. Microsoft. [Online]. Available: https://powerbi.microsoft.com/
8. ŘSD ČR, Požadavky na systém značení provozních celků a elektrických zařízení na dálnicích, rychlostních silnicích, tunelech a jiných objektech ve správě Ředitelství silnic a dálnic ČR. ŘSD - provozní úsek GŘ, odbor správy dálnic 10 421, 2006.
9. B. Rosner, Percentage Points for a Generalized ESD Many-Outlier Procedure. Technometrics, 1983.
10. Microsoft, Machine Learning Studio. Microsoft. [Online]. Available: https://azure.microsoft.com/en-us/services/machine-learning/
11. GNU, R language. Free Software Foundation. [Online]. Available: https://www.r-project.org/
12. Coghlan, A Little Book of R For Time Series. Release 0.2, 2017. [Online]. Available:https://media.readthedocs.org/pdf/a-little-book-of-r-for-time-series/latest/a-little-book-of-r-for-time-series.pdf