A New Approach for Outlier Detection in Near Real Time

Julius Eiweck
Faculty of Arts, Environment & Technology
Leeds Metropolitan University
Leeds, Great Britain
j.eiweck8739@student.leedsmet.ac.uk

Colin Pattinson
Faculty of Arts, Environment & Technology
Leeds Metropolitan University
Leeds, Great Britain
c.pattinson@leedsmet.ac.uk

Reinhold Behringer
Faculty of Arts, Environment & Technology
Leeds Metropolitan University
Leeds, Great Britain
r.behringer@leedsmet.ac.uk

Alexander K. Seewald
Seewald Solutions
Leitermayergasse 33/24,
1180 Vienna, Austria
alex@seewald.at

Abstract — **Outlier detection methods have been suggested for a broad range of applications. They are obligatory in different monitoring tasks (e.g. mobile phone monitoring, credit card usage monitoring). Aim is the detection of sudden changes in the usage pattern which may indicate deviations from normal usage. Parametric (statistical) and non parametric methods combined with univariate and multivariate methods form most of the body of research in anomaly detection. In this paper a novel approach for outlier detection is proposed. For the validation of the new approach we have applied the method on voice traffic time-series obtained from a real-life mobile network. Main focus is the incorporation of standard statistical methods, the ability to implement specific heuristics derived from domain expert knowledge and near real-time performance. The reliability achieved with this novel approach for outlier detection is sufficient to be used in operational processes. Enabling the detection of outliers in near real time supports a broad range of monitoring tasks to be performed by the operational staff.**

Keywords-component; outlier detection; time-series; voice traffic; supervised learning; time-series analysis

I. Introduction

Considering the daily tasks of a network operator to be performed in order to ensure the availability and performance of the services for the customer, the question can be asked what information about the network would be interesting to know in real time. Asking different operators one can obtain various answers but it is evident that the most interesting information is the occurrence of problems in the network that degrade the availability of services for customers (e.g. outliers). Current operators are mainly concerned with ensuring high availability of customer services based on reactive maintenance. Practical interest in outlier detection for certain network parameters is therefore effectively quite high.

Outlier detection is a critical task in many environments and can denote anomalous objects in data continuously obtained from a technical system. Suitable methods have been suggested for a broad range of applications. Parametric (statistical) and non parametric methods combined with univariate and multivariate methods form most of the body of relevant research. Due to the nature of voice traffic time-series, outlier detection on these time-series forces the joint application of such different methods. Thus, these approaches require complex models which are not feasible to be implemented in real operational processes of an Information and Communication Technologies (ICT) network. Operational processes in ICT networks require easy to use and fast applications for monitoring observed systems which are suitable to detect certain anomalies in near real time. Particular anomalies (e.g. traffic overload, outages of network components,...) which are expressed by outliers in related time-series are described in III.B.

Our new approach, "OutLier Detection in NEar Real time" (OLDNERT) enables accurate and fast detection of outliers - especially in voice traffic time-series measurements obtained from telecommunication networks - which represent a crucial factor in operational processes. Due to the relatively high costs for reactive maintenance activities in case of false alarms the reliability is also a critical factor. Contrary to other approaches, our method works in near real-time and can be implemented within existing operational systems.

II. Related Work

A broad range of contemporary techniques for outlier detection, their motivation, advantages and disadvantages are given in the survey by [1]. This paper provides a broad sample of current techniques covering statistical, neural and machine learning methods. Finally the authors stated, that the developer of an application for outlier detection should select an approach that is suitable for their data set.

Different approaches to outlier detection in cellular network data exploration are given in [2]. In this paper several methods for analysis of GSM (Global System for Mobile Communications) network data have been studied for outlier detection on different Key Performance Indicators (KPI) obtained from historical data. All methods can find the most severe outliers in the data.

he classic reference on outlier detection methods is given by [3]. The authors define the characteristics of an outlier as "an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism". This definition clearly assumes a comprehensive knowledge about non-suspicious observations. For the interested reader we refer to several examples of published surveys which are concerned with these topics (e.g. [4], [5], [6]).

III. Definitions

A. Observed Network

This paper is focused on the investigation of the usage of a simple voice service provided by a GSM network. The observed network consists of more than 4,300 Base Transceiver Stations (BTS) located in Austria. The BTS is the network element responsible for maintaining the air interface. It has one to several transceiver/receiver (TRX) units with a dedicated radio frequency each.

The time-series, proposed by the performance management system of the network, consist of 3,864 values (TCH_Traffic in ErlangB) which represent the accumulated traffic per BTS in successive hours. Almost each of the collected voice traffic time series explored in this paper has its individual characteristic in terms of measurement profiles over time.

The following Figure 1&2 show a small proportion of the different profiles selected out of more than 4,300 collected profiles: Different amounts of traffic with daily fluctuations, typical weekly (e.g. working days Monday - Friday, weekend Saturday, Sunday in Figure 1) and seasonal variations (e.g. winter season Figure 2) as well as calendar dependent fluctuations and variations (e.g. Christmas holidays observable between hour 1,100 – 1,600 shown in Figure 1) challenge the construction of a feasible and commonly applicable model for the detection of certain anomalies.

A comprehensive overview on how to deal with the seasonal coefficients is given by Chatfield [7], who clearly mentions the complexity of the procedures that employ a series of filters and adopt a recursive approach.

B. Anomalies in the Network

Anomalies in the network are usually expressed by particular values of measurement data (e.g. performance indicators) or messages (e.g. alarms, event logs) derived from the network. From a statistical point of view these values or messages can be identified as outliers of the expected distribution. Usually, outliers found in real measurements badly degrade the results of empirical model identification procedures [10]. In opposition to these procedures, here the main goal is to identify outliers in the data that represent anomalies in the network. The aim is to enable further explorations of dependencies related to the observed system.

Unexpected overloads or outages of specific network elements are typical anomalies. In certain cases these exceptions are identifiable by relatively simple statistical methods. Anomalies arise due to different causes. Mechanical faults, changes in system behaviour, fraudulent behaviour, human error, or instrument error represent typical root causes for outliers. Their detection can identify potentially system faults before they escalate with disastrous consequences. The most important factor in these processes concerns time. The sooner a problem can be detected, the faster it can be reacted to. Ideally, the problem can be resolved before a customer notice it.

For this purpose it is crucial to monitor meaningful parameters which express a broad range of network properties including specific anomalies. Key findings proposed in [8] and [9] concern the ability of the Key Performance Indicator (KPI) *Voice Traffic* (TCH_Traffic) to express a broad range of network properties including certain problems. The characteristics of the corresponding *Voice Traffic* profiles provide meaningful information about the healthy status of the related network components or even about the entire network.

Outliers found in voice traffic data typically represent problems in the network and the availability of the belonging services (e.g. voice service). The main reasons are caused either by unexpected increase of the use of a service or by degradation of related network components due to different impacts. These exceptions are usually identifiable by analysis of corresponding time plots (e.g. Figure 5-7). Based on these particular identified problems, further investigations in terms of root cause analysis may be triggered.

C. Outlier Definition

For the analysis of performance data like voice traffic time-series it is essential to understand the nature of outliers. In principle outliers are qualified as data objects that do not comply with the general behaviour or model of the data [3]. In this paper, particular attention is given to unexpected zero outliers as well as unexpected peaks and valleys.

As an example for randomly occurring outliers, strong thunderstorms with a high frequency of lightning strikes may cause a total breakdown of a base station and its related services. Such an incident could be detected by different methods, either in near real time by recognition of alarm

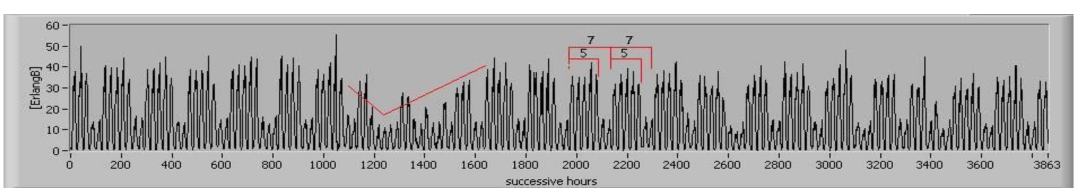

Figure 1: Constant variation in weekday (5 days), weekend, holidays and Christmas time variations

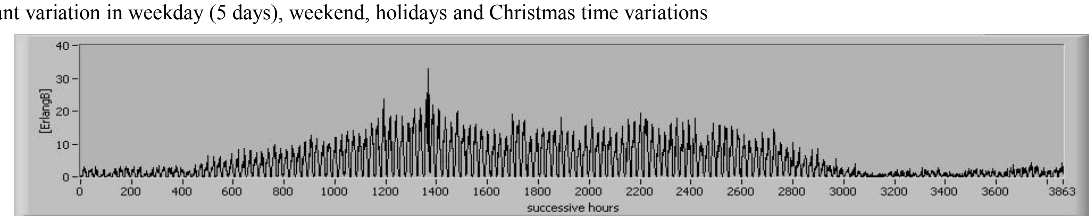

Figure 2: Seasonal (winter) variation with peaks on holidays

bursts generated by connected network elements or by analysis of the performance data where this problem may appear as zero traffic within the corresponding time period (see Figure 5).

Another example depicted in Figure 7 shows some “unexpected” peaks caused by a major event with more than 5,000 visitors present near the related BTS. From an operational point of view such an event might not be unexpected, since event schedules are usually known in advance and could therefore potentially be considered for appropriate preventive actions.

Due to its descriptive nature, the definition of typical outliers in time-series (especially in voice traffic) proposed multiple times in literature (e.g. [3]) is not sufficient for the development of common methods for an automated detection. This section describes a classification of known outliers in order to prevent the learning of other suspect deviations derived from voice traffic time-series.

For the classification of outliers it is essential to know the “normal” variation of the intraday traffic.
Generally four phases (a-d) within the intraday traffic profile shown in Figure 3 are observable:
Figure 4-5 show typical examples of traffic profiles (3 consecutive days) with different suspicious values during the course of the day. Significant and unexpected deviations in the profiles of these time series are evident and may be characterised as potential outliers or anomalies.

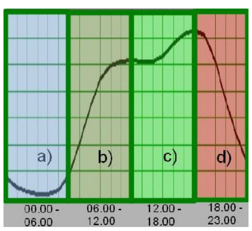

Figure 3: Intraday phases of traffic

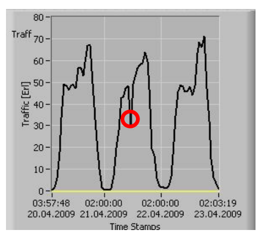

Figure 4: diminished traffic

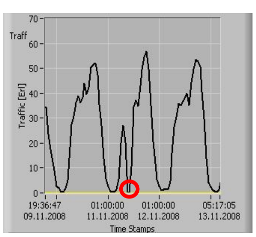

Figure 5: zero traffic

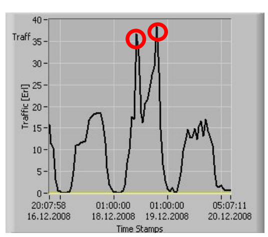

Figure 6: High Traffic peak

Both, unexpected diminished traffic (Figure 4) or zero traffic for a certain time period (Figure 5) typically represent a problem in the access network (e.g. air interface, TRX, signalling channel, etc.) originating in the environment within the related BTS. The root causes for these deviations in the profiles can be very diverse but at least they characterise problems that may be perceived by customers.
High Traffic peaks (Figure 6) which occur unexpectedly (e.g. road traffic jam) or expectedly (e.g. sport events) may indicate an unusual crowd of people within the range of the affected BTS.

In general, the following classes of suspicious deviations from the normal traffic distribution are identifiable:

1) Zero traffic outliers

Outliers of this type are represented by zero traffic in a certain time period during which no zero traffic is usually expected. A typical example is shown in Figure 5. An issue with this method appears especially during intraday phase (a) (00.00-6.00).

2) Positive peak outliers

In opposite to zero traffic outliers, positive peak outliers are classified as unexpected and fast raising traffic deviations with a statistical relevance. The meaning of the term “unexpected” has to be put into perspective in case of correlation with expected major events or seasonally reoccurring situations like New Year's Eve.

3) Negative peak outliers

Significant and fast decrease of traffic during intraday phases (b), (c) and (d) (see Figure 3) without zero traffic could also be detected by standard statistical methods like positive peak outliers. Compared with positive peak outliers no reiterative decreased traffic (>3 hours) has been observed.

In some literature zero traffic outliers and negative peak outliers are described as low-scale event and positive peak outliers as large-scale events [9]. From a statistical point of view outliers are also called additive or innovation outliers. Additive outliers are affecting only a single observation and can be thought as the original observation plus/minus a non recurring “blip”. Contrary to this, innovation outliers occur in noisy processes and also affect the subsequent observations [7].

IV. New Approach

The new approach, “OutLier Detection in NEar Real time” (OLDNERT) aims to enable accurate and fast detection of outliers. The method has been applied on voice traffic measurements (time-series) obtained from telecommunication networks. Contrary to most other approaches, our method works in near real-time. The delay for the detection of outliers is only dependent on the time interval for the collection of the data.

To build a model for the detection of anomalies in voice traffic time-series, a broad range of properties have to be considered. For this reason this section describes a novel approach for outlier detection in voice traffic time-series chosen after evaluations of several standard methods already referenced in the sections above.

The main objectives of the modelling process are:

- Development of one model applicable on certain classes of voice traffic time-series
- Incorporation of particular heuristics gained from domain experts
- (Near) Real time capability
- Minimum computational costs

The result is a robust algorithm for detection of specific outliers in voice traffic time-series, feasible to be integrated in operational processes. The new outlier detection process considers four main topics:

- Compensation of daily variation by transformation of the data into normal distributed classes (day class, hour class)
- Modelling and filtering of seasonal effects (optionally)
- Application of simple statistical methods (IQR, Six Sigma) for outlier detection
- Application of predefined heuristics (optionally)

A. Data Transformation

The visual analysis performed on the 3D-plot shown in Figure 7 inspired to investigate the data from another perspective. The nature and periodicity of voice traffic time-series suggest, that the values of particular day classes and same daytime are very similar.
The following Figure 8 depicts this fact clearly:

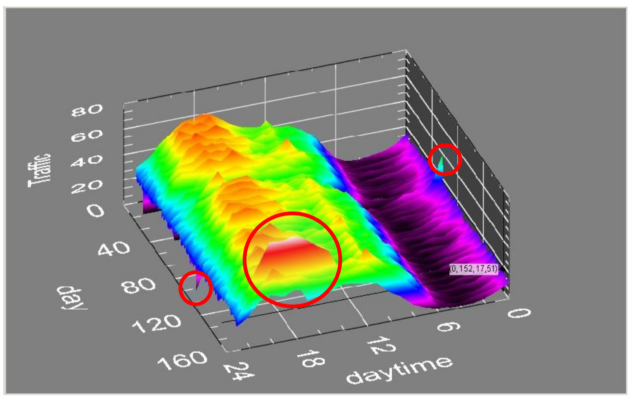

Figure 7: 3D Plot of the voice traffic data (potential outliers marked with circles)

This effect was utilised by incorporating only the values from a certain day time of a certain day class. For this purpose the real values of the training data set were transformed into the matrix $M_{j,k,l}$ by use of a day class vector S_t : where $j=1,\ldots,8$ is the class of the day, $k=1,\ldots,24$ is the corresponding hour and $l=1,\ldots,n$ the number of the values per class and hour.
The day class vector S_t is defined as:

$$S_t=\begin{cases}1-7,\ if\ day=Sunday(1)-Saturday(7)\\ 8,\ if\ day=Holiday\end{cases} \qquad (1)$$

t is the corresponding time stamp of T_t and S_t (successive hours). The Holidays are obtained according to the Austrian Calendar as the voice traffic data was obtained from an Austrian mobile operator.
In addition, the determination of the day classes (1-8) could be performed by a suitable clustering method. In order to improve efficiency of the algorithm the clustering may be performed individually for each BTS

The application of standard statistical methods for outlier detection requires stationary time-series. If strictly periodic variations have been removed and if there is no systematic change in mean (no trend) as well as no systematic change in the variance, a time-series is said to be stationary. These requirements are achievable by transformation of particular parts of non stationary time-series into specific stationary time-series as described above.

In accordance to the similarity effects shown in Figure 8 a time-series $B(l)$ which represents the values for each day of a certain day class and day time can be built by mapping $B_l=M_{j,k,l}$. $B(l)$ usually represents a stationary time-series and is feasible for the identification of potential outliers by simple statistic methods like IQR or six sigma.
The following Figure 9 shows an example $B(l)$

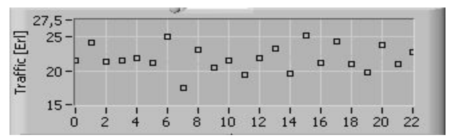

Figure 9: Transformed time-series B(l) with real values for all Tuesdays 8.00 P.M.

where

$$B(l)=M(j,k,l)\ with\ j=3(Tuesday)\ and\ k=20\ (20{:}00) \qquad (2)$$

for a certain TCH_Traffic time-series with moderate seasonal effects.

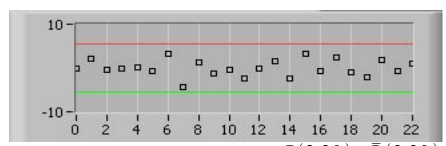
Figure 10: Residuals real values – mean $B(3,20)-\bar{B}(3,20)$ with predefined upper and lower thresholds

As already suspected, $B(l)$ obviously represents a stationary time-series. The outliers are very easily identifiable by simple statistical methods applied on the residuals $B(3,20)-\bar{B}(3,20)$ (See Figure 10).

It is important to note, that this assumption is only valid for TCH_Traffic time-series without significant seasonal variations. How to tackle this problem is described in the next section.

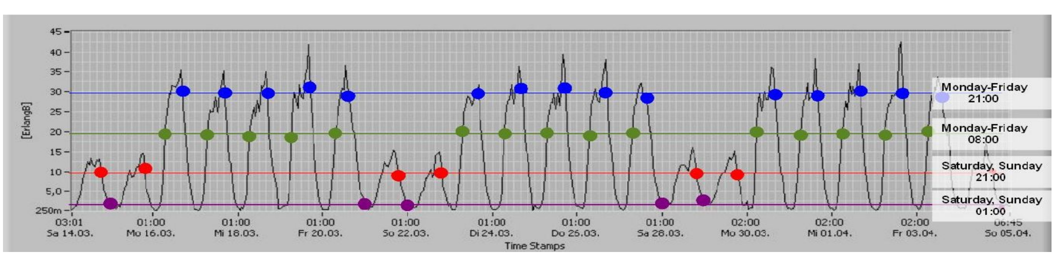

Figure 8: Similar values at certain daytime

B. *Seasonal Effects*

Around 5 percent of the BTS's are located in tourist areas where significant seasonal effects are recognisable. In this cases the time-series becomes nonstationary which prevents the utilisation of simple statistical standard methods for outlier detection.

The transformation of the data eliminates the daily variations as well as fluctuations between successive hours. Only seasonal long term effects are remaining.

An example with strong seasonal effects is shown in the following Figure 11 (real values for all Tuesdays 20:00).

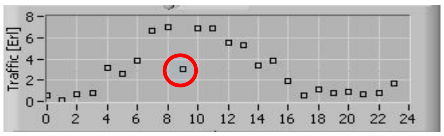

Figure 11: Transformed time-series B(l) with potential outlier

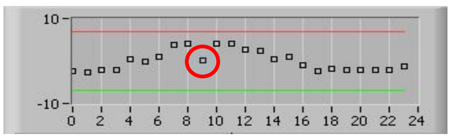

Figure 12: Residuals real values – mean

Using standard statistical methods to calculate the residuals leaves an outlier (n=9) undetected (Figure 12).

Therefore it is necessary to apply a model to filter the seasonal component, which is applicable for all classes and hours. We choose to apply a polynomial regression by calculating $\hat{B}(l)-Y(l)$ where

$$Y(l)=y_i=\sum_{j=1}^{k-1} a_i x_i^j = a_0 + a_1 x_i + a_2 x_i^2 + \ldots + a_{(k-1)} x_i^{(k-1)} \qquad (3)$$

with $k=5$.

Figure 13 shows the polynomial fit with regression order $k=5$ which is sufficient to model the seasonal variations of the specific time period.

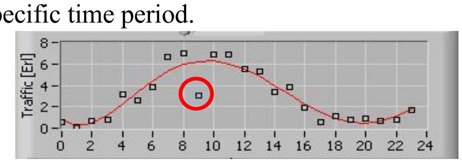

Figure 13: Real values of all Saturdays 19:00. and polynomial fit

Figure 14: Residuals real values – values of polynomial fit $\hat{B}(l)-Y(l)$

In opposite to the example shown in Figure 12 the outlier at (n=9) is now clearly identifiable with standard statistical methods.

C. *Outlier Detection Method*

For experimental purposes the novel model constructed and described in IV. has been applied to the data set of all TCH_Traffic time-series derived from the observed BTS's. The validation of the results for 30 BTS's and particular outliers is provided in section V.

For the identification of potential outliers remaining in the residuals $\hat{B}(l)$ (calculated either by $\hat{B}(l)=B(l)-\bar{B}(l)$ or in case of seasonal effects $\hat{B}(l)=B(l)-Y(l)$) several standard statistical methods are applicable. We have tested two of them – IQR & Six Sigma.

1) *IQR Rules*

The Interquartile Range (IQR) represents the middle 50% of the data which is the difference between the largest third quartile Q3 and the smallest value first quartile Q1. This method is also used for box plots, and is an extremely effective approach, especially when working with large data sets that have continuous data [12].

A common method for outlier detection is the application of the "1.5xIQR" rule which flags potential outliers if the values lie outside of:

upper threshold $Thres_u = Q3 + (1.5 * IQR)$
lower threshold $Thres_l = Q1 - (1.5 * IQR)$

Values in the data set are flagged as problematic outliers if they lie outside of:

upper threshold $Thres_u = Q3 + (3 * IQR)$
lower threshold $Thres_l = Q1 - (3 * IQR)$

2) *Six Sigma*

Six Sigma can be thought as business improvement approach that seeks to find causes of failures in processes focused on outputs that are of critical importance to customers [13],[14]. It assumes that the underlying data values are normally distributed.

In principle, Six Sigma defines a range where the values are classified as normal, and if they are outside of the range they are assumed to be outliers. Usually the range is defined by specification of certain thresholds:

upper threshold $Thres_u = \mu + 3\sigma$
lower threshold $Thres_l = \mu - 3\sigma$

whereby the mean $\mu(B)$ and the standard deviation $\sigma(B)$ are defined as:

$$\mu(B)=\bar{B}=\frac{\sum_{l=1}^{n} B(l)}{n} \qquad (4)$$

$$\text{and} \quad \sigma(B)=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(B(i)-\mu)^2} \qquad (5)$$

and n is the number of available values (e.g. days per day class and day time).

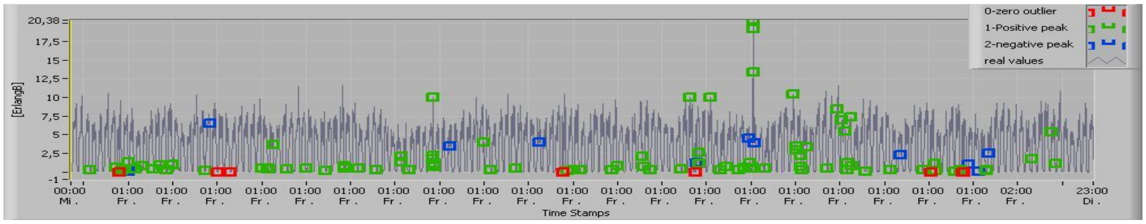

Figure 15: Outliers detected without heuristic rules

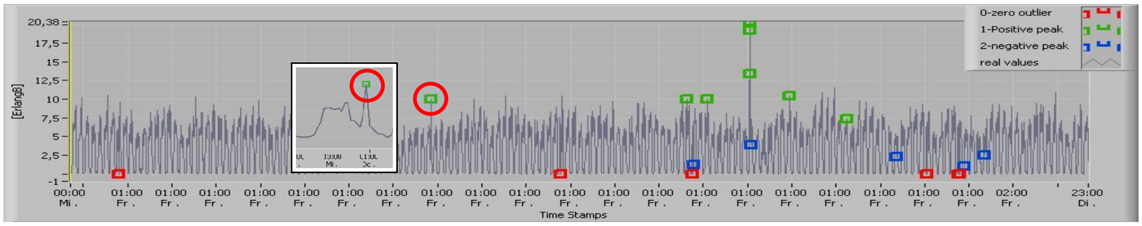

Figure 16: Outliers detected with heuristic rules (The peak marked with a circle occurred on New Years Eve)

D. Heuristic Definition

The application of simple statistical methods for outlier detection in the transformed time-series described above marks all values with statistical significant deviations as outliers. In particular cases (mainly during night time with low traffic) the recognition of these values as outliers is useful to support root cause analysis or other exception handling processes in a real life environment. From a domain expert point of view the remaining outliers have to be filtered by predefined heuristics gained during long term experience in operation of ICT-Networks.

In the following Figure 15 the issues to distinguish outliers caused by a real problem from pure statistical outliers is evident. The standard statistical methods detected 9 zero traffic outliers (red marked), 88 positive peak outliers (green marked) and 12 negative peak outliers (blue marked).

Together with domain experts these results have been assessed. Based on long term experience and comparison with corresponding trouble tickets which are pointing out real problems the following heuristics have been defined for the final construction of a common model:

- Night time (23:00 - 06:00) is excluded from zero outlier detection as traffic is often zero in this period even during normal operation.
- Positive peaks at night time are only recognised if the value is higher than the mean traffic of the affected day (Example of New Years Eve marked with red circle see Figure 16).
- Positive Traffic peaks during daytime (06:00-22:00) are only recognised if peak > 3*daily mean.

Figure 16 depicts again the time-series shown in Figure 15 and the detected outliers (7 zero traffic outliers, 8 positive peak outliers and 5 negative peak outliers) obtained by applying the heuristics described above.

This method reduces the number of detected outliers dramatically and therefore avoids confusion and additional effort for the detailed investigation of irrelevant outliers.

These heuristics are also in line with practically applied rules for measuring the availability of certain services, BTS's and the overall network. It is important to note that each heuristic is strongly dependent on particular operational parameters as well as on business requirements which are individually defined for each specific operator. Conceptual suggestions how to apply the novel approach to outlier detection as well as how to define reasonable heuristics are provided in [8].

Repeated tests have been performed and the results confirmed the reliability of this model. The verification based on the time plots proved the validity of the results. Based on the automatic identification described in IV.C., this analysis has been done for each day class for each hour for all observed days available in the trainings data set.

V. Results

From an operators perspective the most important information gained from the outlier detection process is the occurrence of a zero traffic outlier. It can be assumed that these classes of outliers always point to a problem in the network with impact on end user services.

Therefore, in the first step of the validation process, all TCH_Traffic time-series of the selected sample (30 time-series derived from the previously selected BTS's) have been inspected visually by domain experts. For the period from June 1st 2009 – June 30th 2009 all hourly values from all time-series have been classified for outliers.

Zero traffic outliers with operational relevance were selected and counted based on a predefined heuristic. Finally 22 values out of 21,600 have been identified as zero traffic outliers. This result also agrees with the trouble ticket system which points out the problems related to these outliers.

In the second step of the process the reliability of the novel approach for outlier detection has been assessed. Based on the training data set the hourly data from June 1st 2009 – June 30th 2009 have been classified by the novel outlier detection algorithm according to the outlier definition provided above. For the classification process both IQR and Six Sigma have been used to identify potential outliers based on standard statistical methods. In addition the remaining outliers were filtered by heuristic rules.

Method	Heuristic	True Negative (TN)	False Positive (FP)	False Negative (FN)	True Positive (TP)	Precision	Recall	real outliers	correct identified
		a	b	c	d	P=d/(b+d)	P=d/(c+d)	%	%
IQR	without	21,562	16	0	22	0.58	1	58	100
Six Sigma	without	21,578	0	10	12	1	0.55	100	55
IQR	**with**	**21,578**	**0**	**0**	**22**	**1**	**1**	**100**	**100**
Six Sigma	with	21,578	0	10	12	1	0.55	100	55

Table 1: Confusion Matrices for all applied combinations

The reliability of the classification proposed by the novel approach has been assessed by summarising the results for zero outliers in a confusion matrix [15] (see Table 1).

For the precision P the calculated proportion is 0,58 which means 58% of predicted zero traffic outliers are real outliers. For the recall TP the calculated value is 1 which means 100% of outliers are identified correctly. The following Table 1 shows all confusion matrices as well as their interpretation of the obtained values:
The interpretation of the confusion matrices clearly shows the significant improvement of the reliability of the method obtainable if appropriate heuristics are applied.

VI. Conclusion

In this paper a novel approach for outlier detection on voice traffic time-series has been proposed. Main focus was the incorporation of standard statistical methods and the ability to implement specific heuristics derived from domain experts. With respect to these requirements the developed method is characterised by the utilisation of relatively simple models. In fact the method is applicable to almost the full range of voice traffic time-series obtained from the overall network by minimising computational costs.

The reliability achieved with the novel approach on outlier detection is feasible for use in a real operational processes. In fact, using heuristic rules perfect prediction accuracy is obtained on our real-life test set. Enabling the detection of outliers in near real time could support a broad range of monitoring tasks to be performed by the operational staff.

It is evident that the proposed outlier detection method is also suitable to recognise other threats like intrusion attacks or denial of service attacks. In combination with fraud detection systems based on call detailed records [16] an improvement of service misuse is also achievable. In principle the algorithm is applicable on a broad range of time-series with periodically repeated profiles including arbitrary seasonal effects.

For future research further improvements in the efficiency of operational processes are also expected by combining OLDNERT and appropriate forecasting methods. In particular cases this combination allows the prediction of anomalies in the network with sufficient probability. The knowledge of certain anomalies in advance enables preventive actions by the operator in order to avoid problems before customers notice them.

REFERENCES

[1] V. Hodge and J. Austin, “A Survey of Outlier Detection Methodologies,” *Artificial Intelligence Review*, vol. 22, Oktober. 2004, pp. 85-126.

[2] M. Multanen, K. Raivio, and P. Lehtimäki, “Outlier Detection in Cellular Network Data Exploration,” *Proceedings of the 22nd International Conference on Advanced Information Networking and Applications - Workshops*, IEEE Computer Society, 2008, pp. 1323-1328.

[3] V. Barnett and T. Lewis, *Outliers in Statistical Data*, Chichester: Wiley, 1994.

[4] C. Isaksson and M. Dunham, “A Comparative Study of Outlier Detection Algorithms,” *Machine Learning and Data Mining in Pattern Recognition*, 2009, pp. 440-453.

[5] M. Agyemang, K. Barker, and R. Alhajj, “A comprehensive survey of numeric and symbolic outlier mining techniques,” *Intell. Data Anal.*, vol. 10, 2006, pp. 521-538.

[6] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM Comput. Surv.*, vol. 41, 2009, pp. 1-58.

[7] C. Chatfield, *The Analysis of Time Series: An Introduction, Sixth Edition*, Chapman & Hall/CRC, 2003.

[8] J. Eiweck, “Improvement of Predictive Operational Processes for Telecommunication Networks by Enhanced Data Utilisation,” PhD Thesis in press, Leeds Metropolitan University, 2010.

[9] S.A. Kyriazakos and G.T. Karetsos, *Practical radio resource management in wireless systems*, Artech House, 2004.

[10] R. Pearson, “Outliers in process modeling and identification,” *Control Systems Technology, IEEE Transactions on*, vol. 10, 2002, pp. 55-63.

[11] G. Upton and I.T. Cook, *Introducing statistics*, Oxford University Press, 2001.

[12] D.H. Stamatis, *Six Sigma and Beyond: Statistics and probability*, CRC Press, 2002.

[13] T. Pyzdek, P. Keller, and P.A. Keller, *The Six Sigma Handbook, Third Edition*, McGraw Hill Professional, 2009.

[14] R. Kohavi and F. Provost, “Glossary of Terms,” *Machine Learning*, vol. 30, Feb. 1998, pp. 271-274.

[15] R. Mintram, C. Anyakoha, J. Vincent, and K. Phalp, “Modelling call detail records from a mobile telecommunications network,” *The 16th IASTED International Conference on Applied Simulation and Modelling*, Palma de Mallorca, Spain: ACTA Press, 2007, pp. 56-61.