A new Unsupervised User Profiling Approach for Detecting Toll Fraud in VoIP Networks

Anton Wiens, Torsten Wiens and Michael Massoth
Department of Computer Science
Hochschule Darmstadt - University of Applied Science
Darmstadt, Germany
{anton.wiens | torsten.wiens | michael.massoth}@h-da.de

***Abstract*—Significant amounts of money are lost worldwide due to toll fraud attacks on telecom service providers or their customers. These attacks can be detected or prevented by a fraud detection system. Acquiring labeled data for the analysis of fraud cases is a major problem. This paper proposes an autonomous unsupervised user profiling approach for fraud detection using Call Detail Records (CDR) as data for the analysis and considers problems like random fluctuations in data. Two profiles for each user are used to measure user behavior in different time spans. The two profiles of every user are compared to each other, and changes in user behavior are measured. Describing the change in a numeric value allows checking for extreme changes and detecting fraud. For the detection of random events, a global profile is used. Two profiles are cumulating behavior information for all users, measuring global events in a reliable way. The approach provides low false positive rates. Also, recent fraud cases concerning Fritz!Box Voice over Internet Protocol (VoIP) hardware are analyzed and a detection approach based on this work is proposed.**

Keywords-Call Detail Record; Fraud Detection; autonomous unsupervised user profiling; VoIP.

I. INTRODUCTION

The Internet brought new possibilities for telecommunication (e.g., VoIP), and new communication channels have been created. But fraudsters also found their ways with those new possibilities. Fraudsters invade telephone systems and manipulate them to conduct expensive phone calls at the expense of the owner of the telephone system. The generated cost has to be paid by the users or the service provider most of the time, leading to large amounts of losses and even threatening the existence of small telecom service providers. Telecommunication fraud caused an annual cost in the hundreds of millions EUR at telecom service providers in the last years.

Communications Fraud Control Association (CFCA) reports losses of about 46 billion USD in 2013, an increase by 15% compared to 2011 [1]. But not only cost is a problem caused by fraud. Small providers may also suffer from reputation losses, causing customers to change the provider because of decreased trust and fear of repeated fraud attempts in the future.

The top three methods for telecommunication fraud were Subscription Fraud (subscribing for paid services), Private Branch Exchange (PBX)-Hacking and Identity Theft [1]. The top three types of fraud were Roaming (using stolen access in foreign countries), Wholesale (reselling of stolen user credentials) and Premium Rate Service fraud [1].

The German company “Deutsche Telekom” reported a huge success in the prevention of fraud cases with potential damages of about 200 million Euro, using an automated fraud detection system [2].

Recently, fraud cases were caused by security exploits in AVM Fritz!Box hardware, which is often used in Germany [3]. These fraud cases are analyzed in Section VII, and a detection approach based on the analysis is proposed.

The research project ‘’Trusted Telephony” at Hochschule Darmstadt pursues the goal to increase security and safety in VoIP telephony in cooperation with the German telecom service provider toplink GmbH. A key objective of the project is the development of a fraud detection system, consisting mainly of a software framework.

A huge problem for researching and developing a fraud detection system is the lack of labeled data. In labeled data, each record in the dataset is marked with the appropriate class for the dataset. In toll fraud detection, appropriate classes would be fraud and non-fraud. Labeling requires expertise and is a time consuming process. Because of this, labeled data is often not available, which is why autonomous and unsupervised techniques for fraud detection require less knowledge and personnel to maintain.

For this purpose, a technique has been developed that to work unsupervised and mostly autonomous. Full automation would require a final task for the software, the actual blocking of the customer or destination number. Due to the risk of automatically blocking a non-fraudulent customer or destination number, the approach proposed is autonomous except for this final task, which is done by the system administrator. Unsupervised means in this context that no explicitly generated training data is needed for this technique. It is based on an analysis of Call Detail Records (CDRs) and research on related work and applies user profiling, as well as assorted ideas from related work.

A CDR contains information about telephone calls, e.g., caller and callee, duration, and more. Because labeled data is often scarce, the developed method is designed to work without training a model with labeled data and to autonomously detect fraud in live operation, reducing the need and cost of administration by a staff member of the telecom service provider. The proposed method uses statistical profiles for each user for different time periods and continuously compares them in order to detect anomalies in the users’

 ISBN: 978-1-61208-360-5

behavior. Anomalies are distinguished as extreme changes in user behavior and are used to detect fraud. A Current Behavior Profile (CBP) describes the user behavior in the present, and a Past Behavior Profile (PBP) describes the behavior in the past. The profiles use statistical parameters (features) to describe the behavior in the time span of the profile. With a continuous comparison of those features of both profiles, an estimation of fraud or not fraud is made. This estimation is made by comparing the past profile with the present profile, analyzing extreme changes in behavior.

A. Structure of the paper

In Section II, related work is discussed. A definition of Call Detail Records is described in Section III. In Section IV, the reason for the usage of differential analysis and user profiling is explained as a basis for the concept following in Section V. Section VI describes an experimental evaluation of the proposed method with a first prototype implementation and its results. Finally, Section VII presents a conclusion on the proposed method and gives an outlook on future work.

II. Related Work

In related work, techniques for telecommunication fraud detection that do not require labeled data and are capable of autonomous detection (requiring no administration) are scarce. Much of the related work discusses methods that build profiles from labeled data, train machine learning algorithms and use the result for the evaluation of the data. As mentioned before, expert knowledge and a huge time effort is needed for this task.

Chandola, Banerjee and Kumar present a paper which is rich on information about anomaly detection in general and fraud detection, respectively intrusion detection for telecom networks [4]. As shown therein, most work is based on statistical approaches, neural networks and rule-based strategies.

In [5], two approaches are shown. One utilizes a neural network, trained with profiles and classifying profiles, and the other a statistical approach that has potential for automation. This is detailed more in [6] by the same authors. This method uses two profiles for each user. One is called “current user profile”, the other “user profile history”. The former describes the user behavior in the present, the latter in the past. For the description of the user behavior, so-called prototypes are used to group similar calls by time and duration of the call. Here, a prototype can be seen as a cluster, covering a certain range of values for time and duration of a call. Then, probabilities are calculated for these prototypes using the distribution of calls over the prototypes. A profile consists of probabilities for each prototype. The change in user behavior is measured using the Hellinger distance, which calculates differences between the probability distributions of both profiles.

This technique can potentially run autonomously, but still needs training for the prototypes. Also, the prototypes only use two attributes per call. Adding attributes exponentially increases the number of prototypes. The effects on performance and accuracy for an increased number of attributes are not specified in the paper. The idea to apply two profiles in different time spans to measure changes in user behavior has been a starting point for the method presented in the paper. In this work, the user profiles are built differently, the comparison of the profiles differs as well.

In [7], different user profiles have been evaluated in a combined neural network- and clustering-based technique to detect fraud. One profile type performed better than other profile types and therefore is used in Section V for the profiles describing the behavior of a user in different time spans. The profile consists of the following features: Standard deviation, maximum and mean values for the number of calls, the duration of calls and additionally the maximum cost per call.

In [8], an approach is proposed which combines identity authentication, key process monitoring and anomaly service traffic identification to detect and prevent fraud. There is scarce information on the implementation and no information about the results of the system, e.g., false and true positive rates.

In [9], a more potentially autonomous system for unlabeled data is proposed. It uses a rule-based approach to learn different types of so-called “monitors” that analyze user behavior and alarm the system’s administrator if fraudulent activity is detected. The system still needs templates and learned rules to create monitors, of which the templates need to be prepared and expert knowledge is needed.

Grosser et al. present in [10] an extension of the work in [5] by replacing the prototypes with a self-organizing map. The resulting system still lacks the ability to be autonomous.

In [11], a Bayesian Network is constructed for the detection of fraud in data. It uses the attributes Destination Country, Duration, Call Day and Call Type of a CDR.

As shown in [5], [12] uses a neural network trained with user profiles but different features to classify new profiles with. It results in a true positive rate of 90%, the false positive rate of 10% is quite high.

In [13], different attack patterns and possibilities for their detection are discussed.

In [14], many different approaches are shown: A neural network approach, a Bayesian network and an approach utilizing probability density estimations. All approaches apply user profiles with “…average and the standard deviation of the duration and the number of calls made during the day, maximum duration and number of calls per day during the observed time period…”.

Generally, a lot of work went into the analysis of machine learning techniques requiring training with labeled data which is hard to acquire. Only a few approaches allow to use unlabeled data. Most of them still require some sort of training, making automation hardly possible.

III. Call Detail Records

Each call of customers of toplink is routed through a dedicated voice routing system. Information about the call is recorded as a Call Detail Record (CDR) in text format in a file on a local hard disk drive. The data is then parsed with a parser developed in this project, and the necessary information is loaded into the project’s fraud detection framework. A CDR contains information about the connection and the call, e.g., IP addresses, trunk ID, start time, call duration, calling

 ISBN: 978-1-61208-360-5

number, called number, customer ID, and much more. This data is analyzed for anomalies and potential fraud cases.

IV. ABSOLUTE OR DIFFERENTIAL ANALYSIS

In this section, the concepts of absolute and differential analysis are introduced. An absolute analysis examines a whole set of data, trying to identify fraud cases, but does not consider different types of user behavior. A call that may be treated as a fraud case for one user could be no fraud case for another user. For example, one user only makes long calls to his family at weekends and the other user only makes long calls to his family at workdays. If an absolute analysis considers long calls at workdays as fraud cases, the latter user will be considered as fraudulent, just because his normal behavior does not comply with the definition of normal behavior given by the other user. This problem can be avoided by looking at each user and his behavior differently, thus called differential analysis.

Differential analysis is preferred to absolute analysis in most of the related work, e.g., [5] [7] [10] [14]. The main argument is the ability of differential analysis to include the absolute analysis. In other words, a fraud case detected by an absolute analysis can also be found by a differential analysis, but a fraud case detected by a differential analysis cannot always be found by an absolute analysis [5].

User profiling is a differential analysis method, distinguishing the data by the users in the data. An analysis is then performed for each user on a smaller portion of the data, using only the data of the respective user.

For each user, profiles are constructed to measure the user behavior in a given time span from the user's data. A profile often consists of statistical features describing the user's behavior. For example, the mean duration of all calls or the mean number of calls in a given time span of the user data.

These user profiles are then used for training machine learning or other techniques to detect fraud cases by the values of each profile.

V. BASIC CONCEPT OF USER PROFILING APPROACH

Without labeled data, only few machine learning techniques may be used for fraud detection. Supervised techniques, e.g., a neural network as in [14], need a training phase with prepared, labeled data.

User profiling with statistical methods is therefore used as an unsupervised and autonomous approach. Two user profiles are generated for each user, describing user behavior in two different time spans, allowing for the detection of anomalous changes in user behavior by the comparison of the user's behavior in these two time spans. The user behavior in both profiles is described by the same features. The following sections are giving a more detailed description of the proposed method.

A. Constructing user profiles

For each user, two user profiles exist that represent the present and past behavior in specified time spans. The profile describing the past is called Past Behavior Profile (PBP), and the one describing the present is called Current Behavior Profile (CBP). Each profile uses features, calculated from CDR data, to describe the user behavior in its time span.

1) Features

Features describe different aspects of a user's behavior. In the profiles, the feature vector shown in Table 1 was used:

TABLE I. FEATURE VECTOR USED FOR USER PROFILES [7]

Max Calls	Max Duration	Max Costs	Mean Calls	Mean Duration	Std Calls	Std Duration

These are the maximum values (Max) for calls per hour (Calls), the duration of a call and the cost of a call, the mean value (Mean) and standard deviation (Std) for the same CDR information, except the cost.

For those features, the start-time, duration and cost information of a CDR are needed. The cost of a call is depending on the user agreement and is not given in a CDR. Therefore, an approximation of costs for a CDR was made, based on country code, number type (mobile or fixed-line) and duration.

These features were used because they delivered the best results in [7]. Many works use standard deviation and mean values of the number of calls and the duration of a call to describe the user's behavior. Some works also differentiate them into national, international or mobile [10] [7] [14].

2) Profile Time Span

Each profile P has a length l_P. The PBP additionally has an offset $d \neq 0$, describing the difference in time between the present and the PBP time span (see Figure 1). For a CDR to be included in a profile, it needs to meet the following rules (1) and (2) for the corresponding profile:

$$T_{cdr} < T_n - d \quad (1)$$

$$T_{cdr} \geq T_n - (l_P + d) \quad (2)$$

T_n is the present (n) time, and T_{cdr} is the time of the CDR. If a CDR meets these two rules, it is included in the features of the corresponding profile.

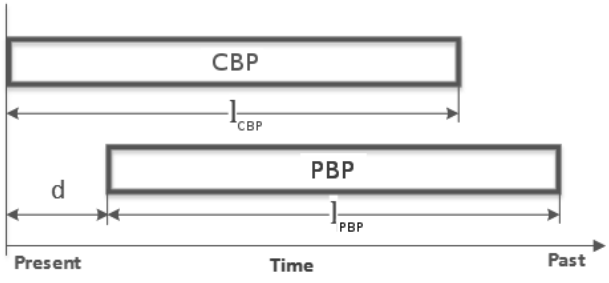

Figure 1. Profile time spans and offset (CBP = Current Behavior profile; PBP = Past Behavior Profile)

The length (time span) of the profiles and the offset are very important parameters for the detection. The longer a profile is, the more CDRs are represented inside a profile and the statistics have more accuracy and less fluctuations. At the

same time, the effects of single fraudulent CDRs become statistically more irrelevant and thus harder to detect. The offset is important for finding fraudulent CDRs that can only be found in groups. It decides how long it takes for a yet undetected fraud CDR to be included in the PBP and therefore make it more unlikely to be found. The length of the offset also affects fluctuations when comparing both profiles. A higher offset causes higher fluctuations, a lower offset causes lower fluctuations likewise.

An optimal tradeoff between the length of the profiles and the offset between profiles needs to be found for best results.

3) Filling Profiles

At first, the profiles need to get filled up for the method to be able to calculate meaningful features. Once the profile contains CDRs for its entire time span, the features can be calculated and used for further analysis. This means that the method has a determined training time for accumulating CDRs that is autonomously done without administration by personnel. In the following, a profile that has been filled up once is called *ready*.

B. Measuring change in user behavior

Once the profiles of a user are *ready*, the change of behavior measured by the profiles can be calculated. This is done by calculating the relative ratio R_F between each feature F of both profiles (PBP and CBP) by (3):

$$\forall F : R_F = \begin{cases} \left(1 - \left(\frac{F_{PBP}}{F_{CBP}}\right)\right), & F_{PBP} \leq F_{CBP} \\ \left(1 - \left(\frac{F_{CBP}}{F_{PBP}}\right)\right), & else \end{cases} \quad (3)$$

This results in a ratio R_F for each feature F, describing the change in behavior for that feature. Each R_F has a range of -1 to 1, with -1 as a maximum decrease and 1 as a maximum increase in behavior measured by that feature.

A ratio R_F for a feature F gives a relative value to the past behavior. It is relative because the severity of a change in user behavior is always relative to the past behavior of the user.

1) Empty profile

In the case that a user did not make calls for a time span greater than the span of all user specific profiles, one of the profiles of a user can run empty. Once a profile is empty, the calculation of the features is not possible, because they attain a value of zero. Comparing a non-empty profile with an empty profile will result in infinite ratios for the features, allowing for detection of fraud where there is none (e.g., when the PBP is empty and the CBP is not empty). Instead of letting the profile run empty, the last CDR in a profile that is about to become empty is not removed. This prevents the features from getting zero values and keeps user specific information for fraud detection. Setting the features to a standard value would disregard user specific behavior and is therefore not done.

2) Features accepting zero

Features like standard deviation can attain a value of zero, even if the profile is not empty. For example, the standard deviation of the duration attains zero, if all calls in the profile have the same duration. Like in an empty profile, zero values are a problem for calculating the ratios. Therefore, a value ε (depending on the range of the specific feature) is added to the affected feature in both profiles.

C. Detecting fraud

For this approach, fraud cases are to be distinguished by extreme changes in user behavior described by each feature. Thus, for each ratio R_F of a feature F, a limit L_F is introduced. Each ratio R_F is therefore checked if its limit L_F is exceeded, and the number (n) of exceeded limits is checked against an additional limit L_E (E for exceedings). If the limit L_E is exceeded, the CDR is labeled as fraudulent and as non-fraudulent otherwise. The procedure can be described as follows:

1. Set $n := 0$
2. $\forall R_F \in R: (R_F > L_F) \rightarrow (n = n + 1)$
3. $result = \begin{cases} fraud, & n > \mathrm{L_E} \\ normal, & else \end{cases}$

Once a CDR in the CBP is labeled as fraudulent, it is to be excluded from inclusion into the PBP. This prevents the PBP from including fraud cases and obscuring potential follow-ups of fraudulent CDRs. This is the first approach chosen for a first experiment. Other approaches for detection using the ratios are discussed in future work.

D. Unexpected fluctuations

Many fluctuations in data and ratios, like weekends and holidays, can be predicted and adjusted for. But there are also fluctuations caused by random events inside the telecom service provider's network, e.g., network, hardware or other failures.

Those fluctuations are hard to predict using user profiles. The idea is to use the relation between absolute and differential analysis. If it is a fluctuation caused by the specific user, the fluctuation is not seen in an absolute analysis. If the fluctuation is global, it will affect all users and will be seen for specific users, too. Therefore, the accumulated behavior of all users has to be measured to detect this kind of fluctuation.

Because the functionality to measure user behavior has already been defined, it can be reused to measure the accumulated user behavior. A global version of a CBP and a PBP is needed for all users. Ratios are calculated the same way as in user profiles. In this case, the ratios are not used for fraud detection, because the source of the fraud cannot be detected by creating profiles for all users. The ratios are used to be included in the user specific ratios for finding the global fluctuations and removing them from user fluctuations.

The inverse ratios of the global profiles are taken to the power of g and are multiplied with the corresponding ratio of a specific user profile as in (4):

$$newratio = (\mathbf{1} - globalratio)^g \cdot userratio \quad (4)$$

An appropriate value for g is determined in Section VI. Both ratios have the same scaling and global ratio that describes the change for the user ratio that is still normal.

 ISBN: 978-1-61208-360-5

Therefore, the inverse is multiplied by the user ratio. Because the global ratio is much more stable with more samples, it is taken to the power of g. g is dependent on the scaling of $globalratio$ and not on $userratio$.

E. Low usage users

An analysis of the data revealed that on average, each user only makes 6-7 outgoing calls per day. About 47% of the users only make 2 calls per day on average. That means a lot of users — and therefore user profiles — include low amounts of calls. Hence, only few samples are available for calculating the statistics, making the statistics inaccurate. A way to handle those fluctuations is to scale the calculated ratios for the user by the number of samples inside the profiles. For the creation of a scaling function$S(x)$, the dependencies of the number of calls in the profiles and the ratios needed to be analyzed. The analysis and the function are described in more detail in Section VI.

Before and after scaling a ratio, it needs to be converted to linear space with (5).

$$S(x,y) = 1 - \frac{1}{\left(\left(\frac{1}{1-y}-1\right)*S(x)\right)+1} \quad (5)$$

x is the number of calls in the PBP, and y is the ratio to be scaled. The part $\left(\frac{1}{1-y}-1\right)$scales the ratio into linear space, and $1-\frac{1}{(\dots)+1}$ reverts it back to the previous space. A full overview of all components and their relationships is shown in Figure 2.

VI. Experimental Results

This section describes the test of a prototype implementation in an experiment. The implementation has been done in Java for an existing fraud detection framework of the research project. The data used for the experiment has been generated by a live environment, recorded by the VoIP switching device. The data consists of 76,326 cost impending calls and spans over a time of one month. It has been anonymized in accordance to the German Federal Law on Data Protection.

For the experiment, the whole data set was used, as the system trains on live data with the assumption that fraud cases are rare enough that the profiles can initially be trained by themselves without greater risks of being manipulated by fraud cases. Assuming the contrary is true and the first data set is containing fraudulent CDRs, the impact would only be that no fraud cases are detected until the fraudulent CDRs are no longer used for the PBP.

For the experiment, profiles of a week's length and with an offset (d) of one day for the PBP are used. In a first run, all occurring ratios are recorded to calculate limits for the ratios, to analyze the parameters for the scaling function and to integrate the global ratios into user profiles. In a second run, the limits were applied and the fraud detection component was enabled.

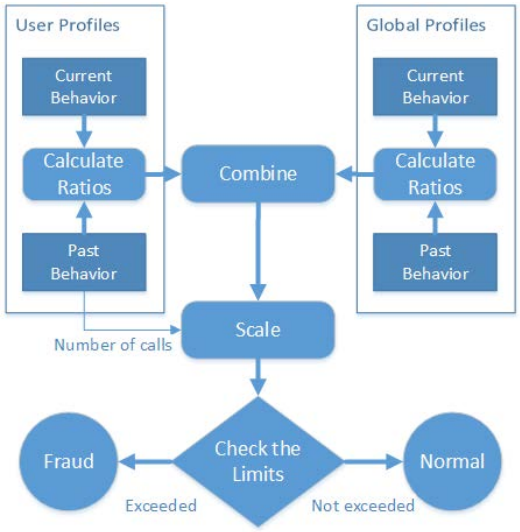

Figure 2. Overview of the components and their relationships

A. First results

For the first results, without incorporating the global profiles and the scaling function, the false positive rates (FPR) for different limits were measured. The false positive rate is a very important measure that indirectly determines the expenses due to inefficiency, because administrators need to look at false positives.

TABLE II. First results of FPR without global profiles and scaling for different limits

Limit for all ratios	Limit for exceedings	FPR
0.25	>0	0.2142
0.25	>1	0.1274
0.5	>0	0.0685
0.5	>1	0.0444
0.75	>0	0.0211
0.75	>1	0.0145

Table II shows empirically tested limits for ratios and the number of exceedings. The FPR has been measured from 50,893 samples, where the profiles were *ready*. The limits and the resulting FPRs will be used for comparison with results of the incorporations of global profiles and the scaling function for low usage.

B. Global profiles

For the global profiles, the same length and offset was used, because the ratios can be compared better if the parameters are similar. The number of calls was used as the only feature for the global profiles. For the parameter g for scaling the global ratio, see (4), a test value of 1 was used.

Figure 3 shows the ratios measured for the given data, chronologically sorted. It shows negative ratios during the Christmas holidays in Germany, successfully measuring its effects on the ratios and it can be used to remove those effects from single user behavior. Also, this figure shows when the profiles became *ready*.

 ISBN: 978-1-61208-360-5

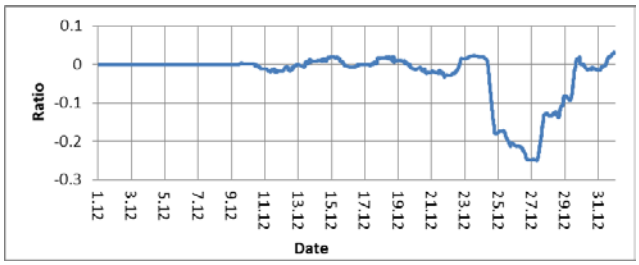

Figure 3. Ratios for number of calls for the whole data in global profiles

The incorporation into profiles of a week's length showed no significant improvements in the FPRs. On the other hand, a small scale test of profiles with a day's length showed very good results in removing weekend fluctuations from the profiles. Figure 4 depicts an example for day-length profiles.

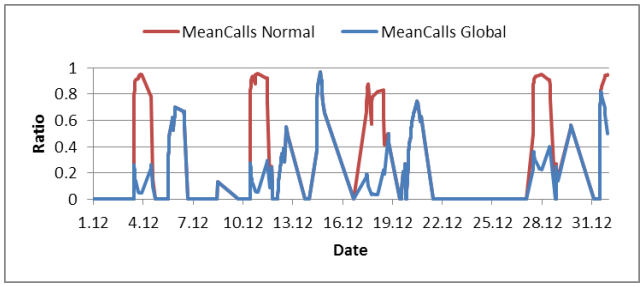

Figure 4. Example incorporation of global ratio into a day length user profile for feature MeanCalls

The figure shows two curves, MeanCalls Normal showing the ratios of the feature MeanCalls without correction by global profiles and MeanCalls Global with correction by global profiles.

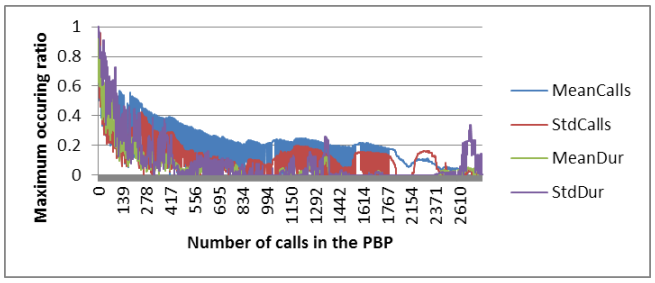

Figure 5. Example for the dependency of max values of the features MeanCalls, StdCalls, MeanDur and StdDur to the number of calls

C. Scaling for low usage

To find an appropriate scaling function, the dependency of the number of calls to the maximum occurring ratios was analyzed. Figure 5 shows an example for four features. It depicts how a low number of samples/calls in a profile can affect the ratios. Therefore, a scaling function was created that scaled the ratios from 0 to 70 calls.

For the scaling function, a simple parable of the form $y = (ax)^2 + b$ was chosen after testing different curves, because it corresponds well to the curve in Figure 5. Using the coefficients $a = \frac{1}{67.1}$ and $b = 0.2$, the scaling begins at 0.2 with 0 calls and ends at 1 with 60 calls with a slight increase. Because about 47% of users only conduct about two calls per day, the scaling function greatly improved the FPRs, as shown in Table III.

TABLE III. CHANGES IN FPR WITH INCORPORATION OF THE SCALING FUNCTION

Limit for all ratios	Limit for exceedings	Old FPR	New FPR	Change in %
0.25	>0	0.2139	0.1684	-21,27%
0.25	>1	0.1272	0.0939	-26,17%
0.5	>0	0.0683	0.0491	-28,11%
0.5	>1	0.0443	0.0290	-34,53%
0.75	>0	0.0211	0.0136	-35,54%
0.75	>1	0.0145	0.0083	-42,75%

D. Determination of limits

The best way to determine the limits is to optimize the ratio of true positive rate to false positive rate. However, this requires labeled data to be possible. Because of the lack of labeled data, the limits were determined by measuring the 99.5% quantile of all occurring ratios for each feature. The ratios are presented in Table IV. Using these limits, the measured FPR is 1.87%.

TABLE IV. LIMITS FOR FEATURES (99.5% QUANTILE)

Feature	Limit
MaxCalls	0.8247
MaxDur	0.6692
MeanCalls	0.7512
StdCalls	0.8270
MeanDur	0.2985
StdDur	0.5400
MaxCost	0.7387
Mean	0.3835

E. Results

Of the 50,893 analyzed cost impending calls, 1.87% were measured as false positives. Through empirical inspection of the false positives, two users were found with an exceptionally strange behavior pattern. The duration of calls and the number of calls per second was the same in about 200 calls, which is very suspicious. After consultation with toplink GmbH, those calls were considered fraud cases. This shows that the presented approach can detect false positives and reduce the FPR to 1.22%, but does not provide a true positive rate for a decent comparison with related work. Still 90.23% of the fraudulent calls found in these two users were marked as fraud by the proposed approach. Compared to the approach proposed in [6], which also proposes a statistical, unsupervised method, the approach of this paper has a lower FPR (1.22% to 4.0%). Compared to other supervised techniques, like [12] (with 50% TPR and 0.3% FPR) or [14] (two approaches with 70% and 80% TPR and 0% FPR for both), the proposed approach has a good TPR and FPR and needs no effort for preparing supervised training data.

VII. CONCLUSION AND FUTURE WORK

This approach allows the detection of fraud cases using unlabeled data and needs no maintenance by an administrator

concerning data for training. Only administration for a final decision on positively identified fraud cases is needed. It is not complex and highly modifiable. It has a low false positive rate and allows detection of fraud cases with an estimated high true positive rate. The scaling for users with low usage rates still needs adjustment, and more profiles need to be tested with other features.

In the future, an autonomous limit adaptation is scheduled to be developed, making manual calculation of limits for the ratios obsolete, and making this approach even more autonomous and efficient. Because of the adapted limits, scaling the ratios for users with low activity is not needed anymore. Also, the limits will provide a more stable FPR for seasonal and user dependent behavior changes.

Furthermore, automation of unsupervised techniques requiring training could be possible by using a sliding window approach on the data consisting of present and past profile values used for training and testing. A support vector machine (SVM) is foreseen to be utilized, possibly including a feature preparation method, as proposed in [15]. This could also be seen as a test for using the results of the proposed approach as an input for supervised techniques. As mentioned in Section II, only few works are available that use techniques capable of being unsupervised and autonomous. Most approaches use techniques requiring training with labeled data and have no potential for automation.

Recent fraud cases, allowed by security exploits in Fritz!Box hardware, showed a repeating pattern in fraud attacks. These attacks utilized the hardware of many customers to call a single fee-based service or number, obscuring the attack by generating only few calls from each customer. A custom version of the approach proposed in this paper will be able to detect such attacks by profiling not the customers, but the destination of the calls. Such a profiling would record the amount of call attempts by different customers to a specific destination and detect extreme changes, enabling detection of fraud cases.

ACKNOWLEDGMENT

The state of Hessen, Germany is supporting this work with funds from the development program LOEWE. toplink GmbH from Darmstadt, Germany is cooperating with Hochschule Darmstadt on this project, providing essential information for fraud detection, e.g., the test data mentioned herein.

REFERENCES

[1] Communications Fraud Control Association, "Global Fraud Loss Survey," October 2013. [Online]. Available: http://www.cfca.org/pdf/survey/CFCA2013GlobalFraudLossSurvey-pressrelease.pdf. [Accessed 23 04 2014].

[2] heise online, "Bericht: Deutsche Telekom wertet Verbindungsdaten sämtlicher Telefonate aus," 10 08 2013. [Online]. Available: http://www.heise.de/newsticker/meldung/Bericht-Deutsche-Telekom-wertet-Verbindungsdaten-saemtlicher-Telefonate-aus-1933436.html. [Accessed 23 04 2014].

[3] AVM GmbH, 06 02 2014. [Online]. Available: https://www.avm.de/de/News/artikel/2014/sicherheitshinweis_telefonmissbrauch.html. [Accessed 23 04 2014].

[4] V. Chandola, A. Banerjee and V. Kumar, "Anomaly detection: A survey," ACM Comput. Surv., vol. 41, no. 3, p. 15:1-15:58, 2009.

[5] P. Burge, J. Shawe-Taylor, C. Cooke, Y. Moreau, B. Preneel and C. Stoermann, "Fraud detection and management in mobile telecommunications networks," in European Conference on Security and Detection, 1997, pp. 91-96.

[6] P. Burge and J. Shawe-Taylor, "Detecting Cellular Fraud Using Adaptive Prototypes," in Proceedings AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management, AAAI Press, 1997, pp. 9-13.

[7] C. S. Hilas and P. A. Mastorocostas, "An application of supervised and unsupervised learning approaches to telecommunications fraud detection," Knowledge-Based Systems, vol. 21, no. 7, pp. 721-726, 2008.

[8] Dai Fei Guo, Ai-Fen Sui and Lei Shi, "Billing attack detection and prevention in mobile communication network," in IEEE 13th International Conference on Communication Technology, 2011, pp. 687-691.

[9] T. Fawcett and F. Provost, "Adaptive Fraud Detection," Data Mining and Knowledge Discovery, vol. 1, no. 3, pp. 291-316, 1997.

[10] H. Grosser, P. Britos and R. García-Martínez, "Detecting fraud in mobile telephony using neural networks," in Proceedings of the 18th international conference on Innovations in Applied Artificial Intelligence, Bari, Italy, Springer-Verlag, 2005, pp. 613-615.

[11] T. Kapourniotis, T. Dagiuklas, G. Polyzos and P. Alefragkis, "Scam and fraud detection in VoIP Networks: Analysis and countermeasures using user profiling," in 50th FITCE Congress , 2011, pp. 1-5.

[12] Y. Moreau, H. Verrelst and J. Vandewalle, "Detection of Mobile Phone Fraud Using Supervised Neural Networks: A First Prototype," in Proceedings of the 7th International Conference on Artificial Neural Networks, Springer-Verlag, 1997, pp. 1065-1070.

[13] M. Nassar, S. Niccolini, R. State and T. Ewald, "Holistic VoIP intrusion detection and prevention system," in Proceedings of the 1st international conference on Principles, systems and applications of IP telecommunications, New York City, New York, ACM, 2007, pp. 1-9.

[14] M. Taniguchi, M. Haft, J. Hollmen and V. Tresp, "Fraud detection in communication networks using neural and probabilistic methods," in Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, 1998, pp. 1241-1244.

[15] D. Wang, Q.-y. Wang, S.-y. Zhan, F.-x. Li and D.-z. Wang, "A feature extraction method for fraud detection in mobile communication networks," in Fifth World Congress on Intelligent Control and Automation, vol. 2, 2004, pp. 1853-1856.

 ISBN: 978-1-61208-360-5