Applying intelligent statistical methods on biometric systems

Willie Betschart

Degree of Master of Science in Electrical Engineering

Supervisors: Per Cornelius
Department of Telecommunication and Signal Processing
Blekinge Institute of Technology
Babak Goudarzi Pour, Optimum Biometric Labs, Karlskrona

Abstract

This master's thesis work was performed at Optimum Biometric Labs, OBL, located in Karlskrona, Sweden. Optimum Biometric Labs perform independent scenario evaluations to companies who develop biometric devices. The company has a product Optimum preCon^{TM} which is surveillance and diagnosis tool for biometric systems. This thesis work's objective was to develop a conceptual model and implement it as an additional layer above the biometric layer with intelligence about the biometric users. The layer is influenced by the general procedure of biometrics in a multimodal behavioural way. It is working in an unsupervised way and performs in an unsupervised manner.

While biometric systems are increasingly adopted the technologies have some inherent problems such as false match and false non-match. In practice, a rejected user can not be interpreted as an impostor since the user simply might have problems using his/her biometric feature. The proposed methods in this project are dealing with these problems when analysing biometric usage in runtime. Another fact which may give rise to false rejections is template aging; a phenomenon where the enrolled user's template is too old compared towards the user's current biometric feature. A theoretical approach of template aging was known; however since the analysis of template aging detection was correlated with potential system flaws such as device defects or human generated risks such as impostor attacks this task would become difficult to solve in an unsupervised system but when ignoring the definition of template aging, the detection of similar effects was possible. One of the objectives of this project was to detect template aging in a predictive sense; this task failed to be carried out because the absence of basis performing this kind of tasks.

The developed program performs abnormality detection at each incoming event from a biometric system. Each verification attempt is assumed to be from a genuine user unless any deviation according to the user's history is found, an abnormality. The possibility of an impostor attack depends on the degree of the abnormality. The application makes relative decisions between fraud possibilities or if genuine user was the source of what caused the deviations. This is presented as an alarm with the degree of impostor possibility.

This intelligent layer has increased Optimum preCon´s capacity as a surveillance tool for biometrics. This product is an efficient complement to biometric systems in a steady up-going worldwide market.

Keywords: Biometrics, Template aging, Impostor detection, False rejections, Verification results.

Foreword

We are living in a time where the desire for security is rapidly growing. Many governments and companies make vast investments in biometrics as a powerful security and surveillance solution. Yesterday’s science fiction is today reality for many of us. The biometric market is exponentially growing. Biometrics have become better and less expensive and have spread from development desks and high security military buildings into our daily lives in form of login devices on personal computers among other things. The researchers constantly refine existing biometric algorithms and develop new ones based on unique characteristics. The most commonly used biometric is by far the fingerprint, which has a long history especially in crime scene investigations. Face, iris and retina readers are becoming more common. The United States enrols every foreign person arriving at the airport making customs able to discover known terrorists. A world wide decision has also been made to make travellers have passports that are readable to a machine and which includes biometrics. This confirms that the traveller is who he claims to be and that the passport is authentic.

Biometrics has got a promising future. Biometrics by themselves are not absolutely optimal solutions. They should be seen as a very powerful complement to other trespassing securities such as code locks. Which one is the superior can be discussed, but biometrics solutions are able to link the identity to an individual. However, what is built by humans can be destroyed or circumvent by human. The development of people wanting to circumvent biometrics has also gone far. There is also a need for feedback from biometrics if they make correct decisions.

Acknowledgements

I want to thank Optimum Biometric Labs for my inspiring time at their company. I especially want to thank Babak Goudarzi Pour as my supervisor who has shared with me much of his time and experience. Being at a small hardworking developing company with a lot of courage has made me learn a lot besides this project. Those things you learn the hard way and do not find in school books. These experiences have been invaluable. I also gratefully want to thank my supervisor from BTH, Per Cornelius for teaching and giving me good advices. My project would not successfully succeed if I had been working all by myself. Everyone involved is a great part of this.

Grateful thanks to all of you

Table of Contents

Chapter 1

Introduction

This chapter gives an introduction to the thesis work. The project was carried out at Optimum Biometric Labs, OBL, a company working in the fields of biometrics. This chapter describes the problems OBL wanted to solve. It starts with a background description about OBL and their areas of interests. The project objectives will also be described in this chapter.

1.1 Background

OBL's, Optimum preCon^{TM} is a predictive condition monitoring tool for biometric based solutions. OBL want to include an intelligent tool which can do qualified estimated assumptions and predictions for system performance. This system should be capable of detecting rejections caused by impostors and false acceptance when logging on the biometric device. OBL wants to discover effects of template aging so the system administrator can receive early warnings that a person may get problems login on the device. Hence the system administrator can tell a user to update his template before annoying problems starts to occur. Template aging effect decreases system performance in biometric systems [3] and it may be necessary to reenrol, i.e. record a new image or template of a person.

1.1.1 Testing biometrics

When testing biometric devices there are three stages suggested [1]. These are guidelines how to plan and perform quality measures and evaluation. The stages are called:

1. **Technology evaluation**
2. **Scenario evaluation**
3. **Operational evaluation**

Technology evaluation is testing competing biometric algorithms from a single technology. These tests are tested on offline databases where samples have been collected from environments where the biometrics have a fairly chance to operate correctly.

Scenario evaluation is testing performance of a biometric system in a controlled manner with an attempt to approximate a real life scenario.

Operational evaluation is testing the performance of a complete biometric system in a specific application environment with a specific target population. Many biometrics are very environment sensitive, for example face recognition readers are light sensitive. A certain biometric device may have different performance results depending on the population type. Gender distribution may also impact the performance. This project is focused on operational and scenario evaluation.

1.1.2 A perfect biometric world

The performance of an optimal biometric system is that all genuine users always pass and all impostors always get rejected. This is not the real word case. There are many factors which may break this desired assumption.

1.2 Project overview

The project overview describes the project objectives and required goals.

1.2.1 Motivation of tasks

As it was described earlier that biometrics are not perfect. There is always a small chance of an impostor being accepted. If the false acceptance rate is high the performance of the biometric system does not live up to the desired security level. Another undesired phenomenon is that the genuine user is being rejected. An overall malfunction list describes these defects:

- **Environmental,** If the environment around the biometric system is not suitable for the devices the error rates may be increased. Some examples of not suitable environment for face recognition are bad lighting, bad adjusted height on camera or a bad distance to object. For fingerprint readers it might be moist, dirt and grease on the touchpad causing worse results. If the devices are outdoors the factors causing higher error rates may just grow in amount.
- **User,** The user himself may cause a long list of problems which may cause defects. User interactions like finger misplacements and wearing glasses may increase error rates.
- **Hardware and software,** these may not perform very good results because of bad algorithms or damaged device.
- **Template aging,** causing more rejections and the user has to be re-enrolled. The template aging effect will be described further on.

The main problem is how one knows if the biometric system does function in an accepted manner in an operational mode. A person who performs impostor-attempts is rejected by the biometrics in most cases but not always. A user with verification problems needs help. An automated system is required to understand these problems and generate alarms whenever needed. The objective of this project is to investigate if neural networks or any other statistical method are proper tools to solve these tasks.

1.2.2 Problems to be solved

A basis for the intelligent layer will be arranged. This desired intelligent tool should be able to discover deviations from the user's normal behaviour. It should be able to understand differences between deviations, like if the user verifies himself badly and is being rejected or if an impostor is trying to verify as someone else. Data should be arranged in a way that template aging effects can be discovered. The application should output an error report and a probability of impostor attack (abnormality). Another problem is to find out if it is possible to predict and detect the template aging problem. Different algorithms should be tested and the system must be able to work unsupervised. This will be done in Matlab. This layer should be working with one-to-one verification.

1.2.3 Problems faced

A major problem was the absence of previously recorded offline data from an operational system. Although, some limited live data were used along with simulated data. The quantised matching distance values from a fingerprint system were analysed. The quantised values were difficult to work with since they vary in a wide range of hundreds to billions. Score is different from each manufacturer and sometimes it is even unavailable. In this project it was necessary to find a way to work with score and find other parameters to work with.

Chapter 2

Feasibility study

A feasibility study is required to understand the nature of the problems presented in this project and also how to develop a solution solving these problems.

2.1 Basic biometric procedure

To understand the definitions and procedures in this report it is important to understand the general procedure of biometrics. An illustration of the general biometric procedure can be viewed in figure 1. A new biometric user must initially be enrolled before using the biometrics. This is usually done by an authorised person who enrols the new user on a biometric device. This authorised person has to control that the enrolee is who he or she claims really to be. When enrolling a template is created and is stored either in a database, a smart card or any other storage medium. If the user is successfully enrolled he or she is authorised to begin with login using the biometrics. The enrolment is done once but later we will see defects which force users to reenrol after some time.

The usage is explained by either verification or identification procedure. These procedures sound very similar but the difference is that verification performs a one-to-one comparison and identification performs one-to-many comparisons. To express this further is for example comparing verification to login on a PC. The user has his user-id and a password. In the verification case the user claims to be himself and log in for example with his fingerprint. Identification is when the biometric unit read a sample and then compare it against a database. This is the procedure when for example identifying criminal suspects. The identification requires longer time than verification because verification just checks against the claimed user's template. In both cases the result is same, either matched or not-matched.

When the device algorithm matches the sample and template a correlation value is calculated. If this value is above a certain threshold value then the individual is matched against the compared template. The correlation value is commonly referred to as score or distance. The threshold value is set after the desired level of security.

The procedure of a verification attempt will further be referred to as a *transaction*.

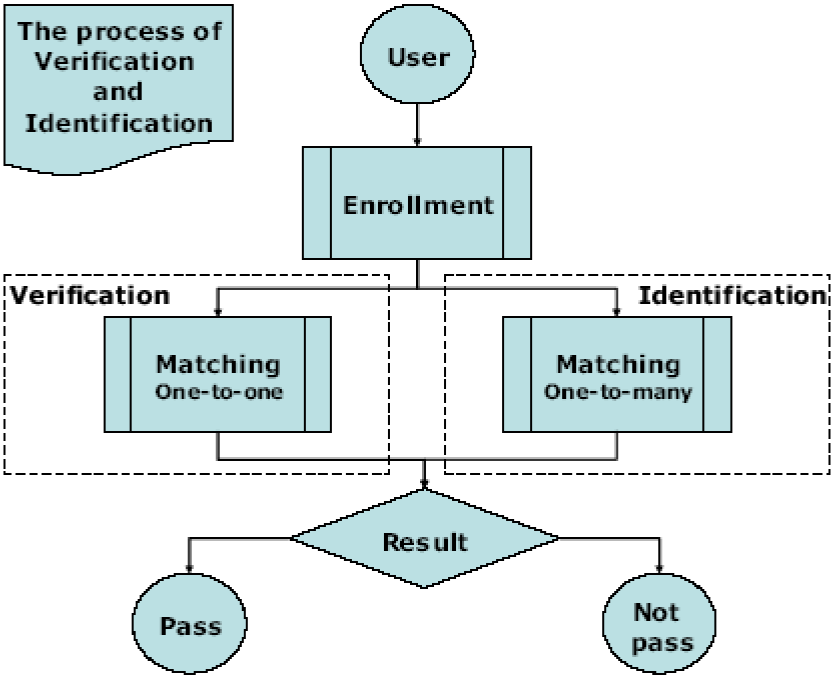


Figure 1: The procedure in a general case of enrolling and matching.

2.2 Definitions in biometrics

The definitions described in this chapter are frequently used parameters in this project.

2.2.1 Biometric algorithm

The biometric algorithm extracts features from the physiological or behavioural characteristics of an individual. It then stores these as a digital signature called a template.

2.2.2 Biometric usage definition

A transaction in this project refers to when an individual makes a verification attempt on a biometric device.

2.2.3 Common parameter definitions

There are frequently used terms in this project which needs a detailed explanation.

- **Score** is a quantized correlation value based on a comparison made by the device matching algorithm where a high value represents a good match between the live sample and the stored template.

- **Distance** is an inverse definition of score. Distance is referred to as a correlated value which should be as near to the template as possible. In this case it represents a low value an accepted value, which has passed the threshold. In this project, values measured are all defined as distance values.

- **Threshold** τ is a value which the score or distance has to pass for the user to be accepted. The threshold controls the security level of the system.

- **Timestamp** is the date and time when the transaction occurred.

Neither score nor distance values are viewed to the user as a verification result. It would not happen because of security reasons. Score is a very secret value hold by the manufacturer. A good reason for keeping this a secret is because if the threshold would be known to users they will soon be hill-climb attacked. Hill-climbing attack is a method used by intruders. If the threshold is known, they retry their attempts to be accepted and hopefully for them increasing their on score value until the threshold are broken.

2.2.5 Error rates

The following measures are calculated in a supervised manner when the objective is to determine the performance of a biometric device. In this project these measures could not be calculated since each transaction occurs in an unsupervised mode.

False rejection occurs when a genuine user is being falsely rejected.

False acceptance occurs when an impostor is being falsely accepted.

False Rejection Rate (**FRR*i***) is an average of number of falsely rejected transactions.
If ***n*** is a transaction and x(n) is the verification result where 1 is falsely rejected and 0 is accepted and ***N*** is the total number of transactions then the personal False Rejection Rate for user ***i*** is

$$FRR_i = \frac{1}{N}\sum_{n=1}^{N} x(n) \qquad (2.1)$$

False Acceptance rate (**FAR*i***) is an average of number of falsely accepted transactions.
If ***n*** is a transaction and x(n) is the verification result where 1 is a falsely accepted transaction and 0 is genuinely accepted transaction and ***N*** is the total number of transactions then the personal False Acceptance Rate for user ***i*** is

$$FAR_i = \frac{1}{N}\sum_{n=1}^{N} x(n) \qquad (2.2)$$

Both FRR***i*** and FAR***i*** are usually calculated as averages over an entire population in a test.
If P is the size of populations then these averages are

$$FRR = \frac{1}{P}\sum_{i}^{P} FRR_i \qquad (2.3)$$

$$FAR = \frac{1}{P}\sum_{i}^{P} FAR_i \qquad (2.4)$$

Equal Error Rate (**EER**), is an intersection where FAR and FRR are equal at an optimal threshold value. This threshold value shows where the system performs at its best.

2.2.6 User characteristics

Biometric users have different characteristics [10]

- Sheep - A user who usually get accepted when verifying.
- Goat - A user who may have problems when verifying.
- Lamb - A user who is vulnerable to impostor-attacks.
- Wolf - A user who performs impostor-attacks.

A so called sheep is a person who the biometrics works pretty well for. The user get good score values and is seldom falsely rejected. It does not work as well for a goat. There are days when a goat can have big verification problems and days when all goes fine. Nothing has to be unordinary about their fingers but it just does not work as well for them. There are actually people who impossibly can use biometrics too.

2.2.7 User distributions

When biometrics is used and distances are generated they get different values *D* according to a probability density function [2]. Genuine users and impostors have different functions. A genuine user has distribution function Ψ_G and impostors distribution function Ψ_I which is illustrated in the figure below. Both functions are approximately Rayleigh distributed but has different standard deviation s. A Rayleigh distribution is defined

$$P(r) = \frac{re^{-r^2/2s^2}}{s^2} \qquad (2.5)$$

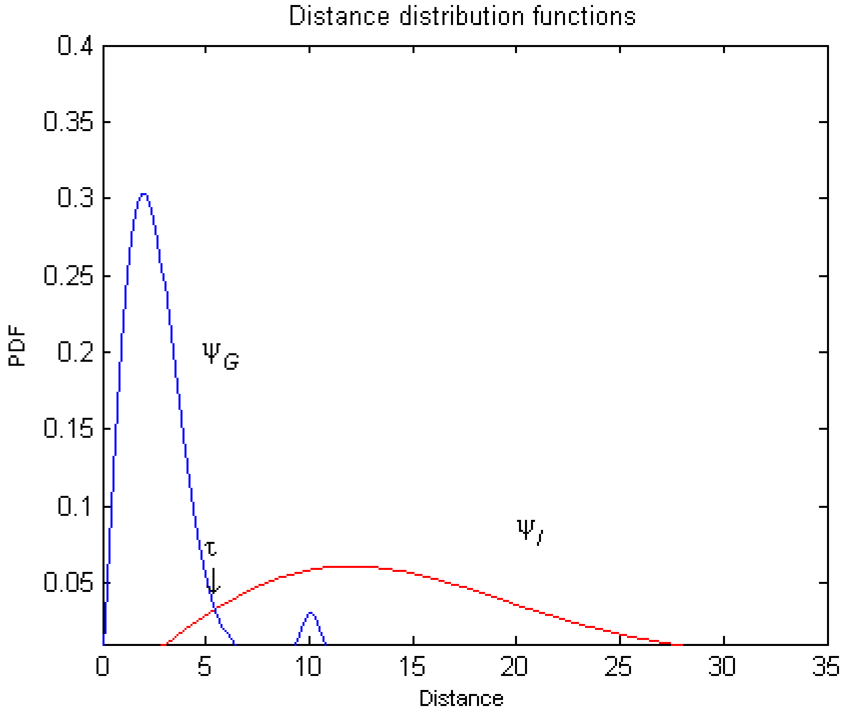


Figure 2: Genuine and impostor distribution functions. This is a symbolic illustration. Distances smaller or equal to the threshold τ are accepted by the system. The small blue part seen at distance 10 belongs to a category of people who has sporadic verification-problems, “goats”.

The distribution functions are estimated over large populations which make this a general approximated model. Every individual user has got own characteristics, but they do not have a completely different model.

2.2.8 Matching error rates

This chapter's equations are based on systems which returns distance values *D*.

The single comparison False Match Rate FMR
FMR(τ) describes the probability for an impostor being able to be falsely matched with a stored template in database. If τ is increased the security is reduced and chances of more false matched impostors increases.

$$FMR(\tau) = \int_0^{\tau} \Psi_I(D)\partial D \qquad (2.6)$$

The single comparison False Non-Match Rate FNMR
FNMR(τ) describes the probability for a genuine user not being matched against his own template. The expression 2.7 says that an increased threshold would increase the area and less false non-match transactions would occur.

$$FNMR(\tau) = 1 - \int_0^{\tau} \Psi_G(D)\partial D \qquad (2.7)$$

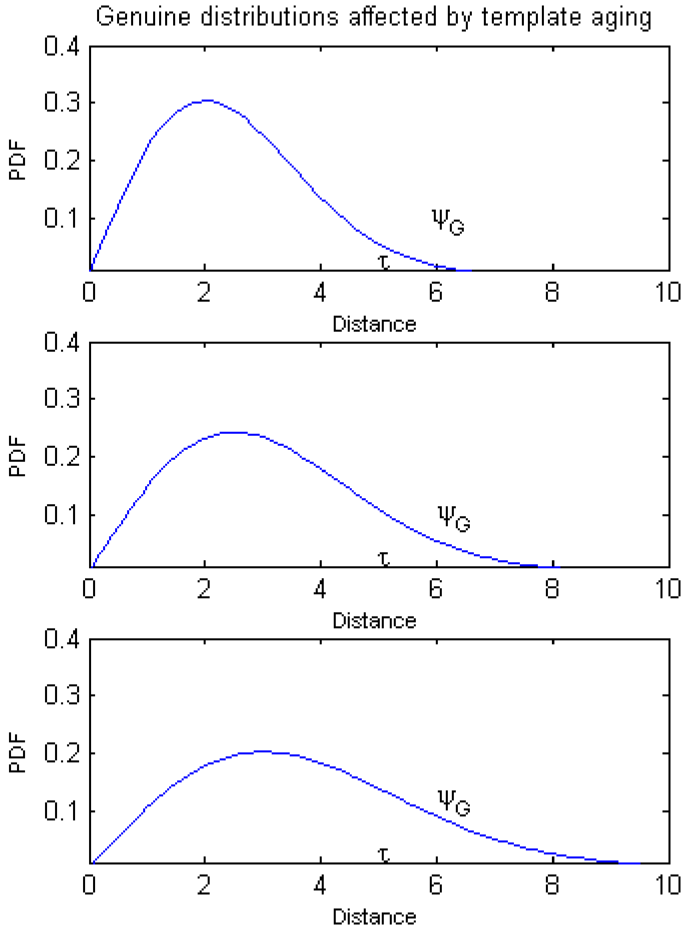
2.2.9 Template aging description

Template aging is a phenomenon within biometrics when the user appearance has changed compared to the stored template in the database. It can depend on scars, growth, age, traumas, plastic surgery, etc. The most vulnerable group using biometrics is children who grow fast and often change appearance. They also hurt themselves when playing. Biometric techniques which are more vulnerable to template aging than others are facial and voice recognition. The human face can change quite often because of many different factors like for instance; beard, make up, sun tan, spots and wrinkles. People can get thicker or skinnier under a short time too. The voice can also easy become different than it is used to be. Then people get a cold, the voice often get hoarse and sometimes unable to speak at all. A boy's voice becomes darker in puberty. The list of factors can grow enormous. Reading the eyes retina is also a known biometric technique. Diseases like glaucoma and diabetes harms the retina in a way that the reader may have problems to verify or identify the user.

Biometric methods which are not equally vulnerable to template aging are fingerprint and hand or palm and iris readers. Especially the iris is very stable throughout the whole life. It is randomly created in the beginning of life and remains that way until the end. An adult's fingerprints do not change much either, but as always injuries like papercuts easily occur.

Template aging is a problem today. Biometric systems suffer from many problems already, mostly security matters. If template aging can be circumvent much resources can be spared. It is annoying being rejected when you are trying to verify you as yourself. The common method to avoid the problem is to reenrol, but when? Enrolling each week, month or year? Which period is most efficient? Reports are very various regarding how long time it commonly takes before template aging occurs. A system with enormous numbers of users has to be efficient. The system administrator has other things to do than enrolling people. Template aging is not the first biometric problem to give priority to.
One important thing having in mind is the fact that many people enrol when they are least experienced biometric users. If the template quality is to low it might become a template aging case if the user becomes more experienced verifies against a bad enrolled template.

Modelling template aging mathematically can be expressed as the genuine distribution function Ψ_G is changing over time [2]. The figure 3 is an illustration of this.



(a) This represents the user's PDF the first time after enrolment

(b) This represents the user's PDF a time after (a), the distribution has started to reshape. The non-matching rate is increasing.

(c) This time the PDF has changed a lot and the non-matching rate is much higher than when the user started using the biometrics.

Figure 3: The three pictures models a template aging scenario. Note that it is the same scale as before and this makes the distribution functions look wider than they are.

2.3 Long term prediction using neural networks

OBL wants to investigate if it is possible to forecast template aging cases using artificial neural networks. The idea is to anticipate a time for the user to get an opportunity to reenrol before it is getting to annoying for the template aging victims.

2.3.1 Introduction to artificial neural network

Artificial neural networks are influenced by the brain's neurons and are mathematically expressed as simplified neurons. Neural networks are widely used in computer vision, signal separation and association applications i.e. recognise handwritten text and prediction. There are many fields where neural networks are very useful. The common factor is their ability to learn and recognise patterns.

2.3.2 Neuron model

The simplest neural network is a single neuron. It consists of input signals **x** and a weight matrix **W**, a summation module with output u and an activation function f(u). This activation function doesn't exist on a linear neuron. The choice of f(·) is usually a sigmoid function, hyperbolic tangent for instance. The network may learn nonlinear patterns because of the nonlinear activationfunction.

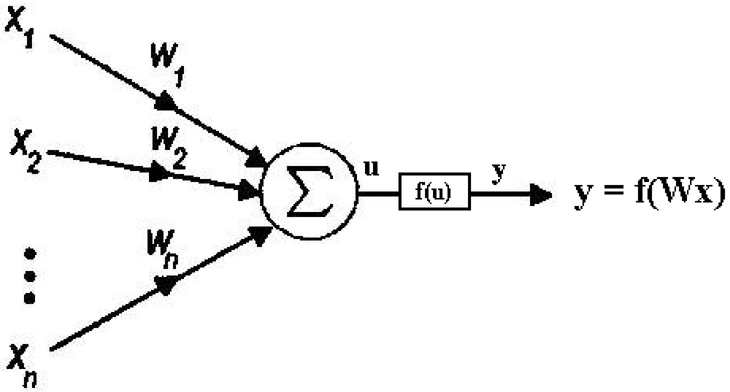


Figure 4: A simple neuron model.

2.3.3 Hidden layers

A common and powerful way building neural networks is to have several layers of neurons. They are called hidden layers. The different layers may have different matrixsizes (amount of neurons) which are chosen depending on the specific data to learn. Using hidden neurons in the network architecture is a good solution when a network should learn difficult nonlinear functions.

2.3.4 Training a network

As mentioned earlier, the point of using neural networks are their ability to learn. There are many different techniques how to do this. There are supervised and unsupervised learning techniques. The supervised requires a target signal, the unsupervised do not.

2.3.5 Supervised learning

Training a single neuron layer can be performed by the ordinary **LMS** algorithm. LMS stands for Least Mean Square and the algorithm minimizes the mean square error. It iteratively calculates the inverse of an autocorrelation matrix. When the gradient error vector has become zero or close to zero the algorithm has converged and found the networks optimum solution.

If there are hidden layers the popular **backpropagation** algorithm is widely used. The backpropagation algorithm has a rather slow convergencespeed hence it has many variations and improvements, like variable steplengths (a.k.a. learning speed), leaky factors (also referred as forgetting factors). Sometimes the whitening technique is used to improve the backpropagation's speed of convergence. If the bowl-shaped error curve is not circular shaped the gradient has got a longer way to its minimum. Whitening is literally finding the principal components and performing coordinate shifting. This reshapes the error bowl into a circular shape and the distance to its minimum solution get closer.

There are also **associative networks** which recognise items from a stored memory.

2.3.6 Unsupervised learning

If no target signal exists, unsupervised learning is often used. There are many existing methods which do this.

Cluster algorithm uses competitive learning. The output from that kind of network has classified the input vector to a certain class or cluster.

Self organized map is a very popular method. The neurons are spread over the signal plane or space and a winning neuron acts as a classifier of the input.

Principal component analysis PCA is a technique how to find principal components or eigenvectors. It is often used in compression and in whitening. The Generalized Hebbian algorithm is a well working way of finding principal components. The principal components are pointing in orthogonal directions.

Independent component analysis ICA is a quite new method and it is commonly used in signal separation applications. Compared to PCA, ICA the components do not need to be orthogonal.

2.3.7 Methods to perform long term prediction

Long term predictions have increased in popularity in many fields. The most known field is prediction of stock markets. There are many people interested if the economy trend is about to change. If it would go down people sell or buy if the trend is going up. Almost the same thing is interesting to look further into in this project. If a user's rejectionrate start increasing there are probably something wrong. Can this possibly be forecasted before it has gone too far?

There have been many who have done predictions in different fields already to observe how the predictions were done. The first thing is to analyse data. If data is very periodically or has correlated dependencies a prediction far ahead is possible. Next step is to manage data in a way to be able to perform the prediction. One of the best examples found was polish scientists who was predicting natural gas load [12]. They have discovered the relationship between amount of gas load and temperature and prepared a proper time coding for their system. Then they put these inputs into a network, building a new time-dependent class of gas load. As a result the comparison between the actual load outcome and the prediction were acceptable similar each other. These guidelines were considered well working when investigating methods to predict biometric data as in this objective.

Chapter 3

Conceptual model

The conceptual model describes solutions to the projects objectives. The first objective is how to perform a long term prediction to be able discovering if the biometric usage performance is decreased in an unacceptable level using neural networks. Abnormality detection system is then described. The direct solutions will not be described here because of confidential matters. The permitted reader is referred to the specified appendix.

3.1 Performing the long term prediction

Optimum preCon^{TM} is an operational level working product and it is therefore very difficult to identify what it was causing the rejections. It can be depending on many facts. The rejections are in sense of being able to determine a pure template aging case caused by the user himself. Impostors who have tried any attempts would weigh into the statistics unnoticed. Therefore a detection of cases similar to template aging is of interest.

Being able to do long term prediction of biometric usage simulations of values are required because of lack of real data. Score can vary in amplitude and can have very low correlation. Scores would be preferred normalized, logarithmic or in per cent. If the intelligent layer would be independent from the biometric layer it must be able to handle all kinds of score values from different manufacturers. Instead the conclusion was to predict rejectionrates instead because the values are independent of different manufacturers and the numeric values are always between 0 and 1. The rejectionrate curve is a trend measure which describes the status of a user over time.

EMA (Exponential smoothed Moving Average) is a well known trend calculation [7], often used in prediction of e.g. market prices. It is easy to see changes. EMA is a moving average for a time period and a bit weight is given to the latest data. The equation is given

$$EMA(t) = \alpha(x(t) - EMA(t-1)) + EMA(t-1) \qquad (3.1)$$

the smoothing constant *α* can be chosen

$$\alpha = \frac{2}{n+1} \qquad (3.2)$$

and *n* is a period for example number of days.

The prediction was performed by EMA trend calculation into x(n) and then put into a tapped delay line with length *p*. The input matrix **x** = **[** x(n-p) … x(n-2), x(n-1), x(n) **]^T** and the target vector d(n) = x(n+1).
A feedforward net was designed

$y_i = f(\mathbf{w}^{[3]}u(\, f(\mathbf{w}^{[2]}g(\, f(\mathbf{w}^{[1]}\mathbf{x}_i)\,))))$ (3.3)

A hyperbolic tangent was the choice of activation function f(·).

The feedforward net was then improved by Jordan recursive network structure.
A Jordan network is similar to a feedforward with the exception of a *context unit* which is an added neuron in the inputlayer. To this context unit is the output y_i recursively an input like an IIR filter.

Training of the network was performed by the backpropagation algorithm.
A variable steplength α was used. α starts high and then decreases α to a very low value.

A one-step prediction was done on each value $\mathbf{x}_i$. A training set was used to train and filter through the network. When the trainingset was out of samples, the output sample y_i were recursively feed into the input signal $\mathbf{x}_i$. This was done for an arbitrary sequence.

A spared set of data was used as validation were the rejectionrate had an up going trend; a simulation of the problems to forecast. When this sequence was not included in the training set it could not been predicted either. The long term prediction resulted completely wrong.

The one-step prediction worked well in the training sequence but when the forecasting of unknown values should be predicted it did not implement good at all. Because of the recursive input it affects the output of the net. It requires being very precise otherwise the error grows exponentially in height. For further reading about the results of this task in the project, see Chapter 4 Unacceptable problem forecast using neural networks.

3.2 Abnormality detection

OBL's surveillance system Optimum preCon^{TM} is enhanced by an intelligent layer which has intelligence about the user's historical behaviour. Being able to discover deviations, information about the past is important. When a user has been active for some time a pattern will be stored in the database in forms of different parameters. Based on earlier behaviours a regular individual pattern is formed and is recalculated to a pattern template. This application will mostly indicate regular use of the system, but if something does not seem to be in order an alarming notification is delivered. If deviations as template aging seem to start to occur, the system is alarming incoming problems. Impostors are difficult to detect unsupervised. One example of scenarios the system is facing is if someone tries to verify himself as another user the biometric device often return a very low score value (see impostor distribution figure 3 chapter 2). An experienced user has passed his learning phase and usually do not receive low scorevalues and a suspicious impostor notification will be delivered. The concept is to determine deviations like the example and make a conclusion if it could have been somebody else than the genuine user.

3.2.1 Impostor detection possibility

All rejected persons by the biometrics are treated as impostors. The biometrics rejects a person if he does not pass the threshold. There are no specific information in the database can tell who it was. A false rejection is quite possible. Studying some supervised samples of impostor attacks was necessary to get information of how parameters would be set to discover a future impostor. The whole perspective was reviewed. It is a complex situation to make a decision to confirm an impostor attack.

Professional impostors, like the ones who use gelatine fingers are hard to detect for the biometric devise [11]. The impostor must exactly know how to lay the artificial finger on the biometric reader to achieve an optimal result even if the artificial finger is almost identical to the real one. Advanced fingerprint readers which measure moist, heat and even haemoglobin flow may also be fooled by these artificial fingers. A good score might be lower than the average of the user, and to decide if it is a fraud or not, the circumstances must be observed to achieve a qualified answer to that question. This is indeed a very sensitive question. The following example illustrates how the system will deal with abnormal situations. A stressed person come to work from cold outdoors late at night because he had forgotten his wallet there earlier. It is very possible he will not achieve a very optimal score. This time it is not regular for the genuine user to login either, even if he tries many unsuccessful attempts. The risk of the system treating him as an impostor is high and that is also the purpose. The deviations from the normal behaviour are too many in this example.

3.2.2 User integrity

Today many people easily get insulted if questions arises doubting their actions. They are also protected by the personal integrity law. Because it is an unsupervised surveillance system users should be informed that if a transaction suspiciously looks like an impostor attack, it is just to guarantee safety that the questions will arise. Learning by real experience will grow the knowledge in this case. The point of Optimum preCon^{TM} is to enhance the personal integrity by being carefully protective and suspicious.

3.2.3 Detailed analysis on statistical properties of current parameters

How is it possible to detect false acceptance or false rejections unsupervised? The biometrics should only reject unpermitted users but they are not ideal. The genuine user might be falsely rejected and if the user has got verification problems, the system should inform about this matter. Does a low score mean that it was a genuine user with an old template or an impostor? Decision making will be performed on parameters which are verification quality measurements. Based on their behaviours they will be compared with verification results using statistical assumptions. This will require highly adjusted parameters and a lot of measured data. This chapter will discuss parameters of interest being able to form a desired decision.

Score values carry a lot of information. One major problem with these values is that they differ with each device. Sometimes score is even not able to achieve from the device. The system should be independent of the choice of biometrics. It has to be adapted to each kind of device used. Is score known to the system then information may be analysed. Score is a more detailed information about the verificationresult than only passed or non-passed.

The **verification result** is a Boolean variable and can only result in accepted or rejected. Verification results have more information if entire events is analysed. A very powerful method is analysing the series of verification results which bring much more information than analysing each single transaction alone.

BIR–Quality is a constant which is a template quality measure. BIR stands for Biometric Identity Record and is a definition in the BioAPI. When a user enrols, a code is stored in database how successful the enrolment was. As mentioned before, users are often not used to biometrics when enrolling. It can be because of the user is unaccustomed with de device that the enrolled template bad. A badly enrolled person has higher probability to get verification problems after some time than others.

Threshold is a parameter which rejects persons who not seems to be the correct user. A threshold set high bring less transactions with false acceptance as results. At the same time more false rejections occur. A low threshold brings higher false acceptance rate and lower false rejectionrate. This could be taken under consideration when making a decision of an abnormal transaction.

The **timestamp** pattern gives many aspects of the users. Especially in interest is when occurrences commonly happen. It could help creating better accuracy in calculations. A high frequent user does not suffer equally much from being rejected as low frequent users, because they are more familiar with the system. An experienced user has a faster learning process how to use the device. An activity status of biometric usage scale is set and varies depending on the aging risk.

3.2.4 Defining abnormality actions from user transactions

The definition of an abnormality is behaviours that differ from the ordinary user's characteristics. The abnormal behaviour is an unexpected action caused by the genuine user. Remember that there are not any observations of the scenarios. The only thing the decision has to rely on is the presented data. This chapter explains suspicious scenarios that this project is about to detect and give an opinion of the degree of impostor possibilities. Simultaneously it is also detect when a user might be needing help because of problems using the biometrics. On the next page some general example of scenarios are illustrated.

Abnormal score scenario 1
A user who normally receives a good score suddenly receives a much worse score value. The score is close to the threshold, characteristics of a false acceptance. The dilemma is if it was a succeeded intrusion or a fumbling caused by the user. The only thing to decide about is if it was an abnormality very suspicious for the current user.

Abnormal score scenario 2
Users who usually do not receive very good score suddenly receive a very good score. This is an abnormality for those users. It might be a good thing for those users but a chance of a well performed intrusion is not to forget.

Users with a more uniformly spread score distribution is impossible to make any decisions about score values. It is normal for them to receive different values.

Abnormal many rejections scenario
A user with a temporary rejectionrate of 33% implies that the user usually fails to verify once every three attempts. This is classified as normal behaviour for that particular user. If more rejections occur the user might have problems verifying and probably needs help. This is a big problem especially if the user never succeeds to pass. If one transaction finally is accepted and the score from that transaction is low then is it even more suspicious.

Single rejected transaction scenario
If a single rejection occurs and it is never followed up by a successful transaction, then it is treated as a very suspicious action caused by the user.

Abnormal timestamp scenario
A transaction which happens on an unusual time is treated as an abnormal behaviour. User's who usually have a biometric usage daytime is it suspicious if a transaction occur at the middle of the night. It is even more abnormal on unusual days, like weekends for an ordinary worker. If schedules can be integrated in the system, recording that the user should actually not be there and that transactions has been made anyway, can higher the risk of fraud attempts.

Combinations of scenarios
A combination of earlier scenarios of suspicious events increases the summation of abnormality rules and also a higher probability that something wrong has happened.

3.2.5 Detecting abnormality actions from user transactions

Each user has their own characteristics when they use biometrics. The patterns are stored as a pattern template and the concept of detecting impostor attempts is influenced by the general biometric procedures. How the template is assembled is explained in *Appendix A - Template Estimation*. This pattern template constitutes the backbone of the whole conceptual model.

The usage of the biometrics is usually done normally without any spectacular events happening and therefore do not these events pass the filter for further procedures. The system is reacting on abnormal events and this procedure is in detail described in *Appendix B- Application structure* and in *Appendix C - Abnormality Detection modules*.

Optimum preCon^{TM} receives messages describing what issued the alarm. The messages contain a degree which describes the possibilities that it was another person who triggered the alarm, depending of the degree of abnormal behaviour. Users who have verification problems get a low rate if the whole perspective does not seem to be too suspicious otherwise they are treated as it were a fraud attempt. This is represented by a high possibility rate which is causing an alarm which should be taken seriously. The solution of how the decision is taken is described in *Appendix E - Transaction Identification, fraud possibility decision making*. This module performs an intelligent decision based on the user's history and the current scenario.

3.2.6 Detection of template aging effects

A true template aging case must be detected under supervision. The actual cause of the decreasing performance must be known. The application is indeed searching for these types of changing behaviour because it can indicate performance degradation. That module is described in *Appendix D - Radical distribution change detection*. This module is interacting with the module in *Appendix E - Transaction Identification, fraud possibility decision making* when performing conclusions in Abnormality detection. Information about the past general usage is very interesting when to decide the degree of impostor possibility. An example to illustrate the importance of this information is if a user who generally never has verificationproblems suddenly appears to have one. This event seems suspiciously to be an impostor attempt. If instead the overall usage has changed to the worse, when it does seem more like that the person has verificationproblems and might need to reenrol, which will circumvent future conflicts using the biometrics.

Chapter 4

Data analysis and results

This chapter show the results from the specified objectives. It begins with data analysis on the assembled data on Optimum preCon^{TM}. The analysis resulted in few parameters to work with. This chapter also describes simulations used to perform the less successful prediction of data, an objective which was cancelled after realisation that it would not work in reality. The simulations were used until enough data was collected. The assumed theory was not that simple when reality came across. The chapter is further describing the abnormality detection results. This was tested in a smaller scenario evaluation during runtime and satisfied results were achieved. The tests showed where calibrations of the algorithms for future works are necessary.

4.1 Data analysis

Data analysis is important to begin with. Processing the data, knowledge about its behavioural characteristics is a fundamental part. The goal is to find patterns in the parameters to use both in finding abnormalities and to predict data. Dependencies in different parameters are necessary in these cases. A working database existed but unfortunately to begin with there was not much previously recorded data to work with. The tasks required data stretching over a longer time period. If the data were recorded quickly, as a high frequent usage in one day, it would give misleading values. The data has to be continuously assembled as an everyday biometric usage to become natural without attempting to manipulate any results. The test population was instructed to perform as good results as ever possible. The population was indeed not very large, only a couple of volunteers. The population included volunteers with different characteristics, which in this case is very good. Almost every volunteer had characteristics which could be described as a sheep and one volunteer had characteristics as a goat. After some time did the one with goat characteristics find it harder to successfully verify and therefore reenrolled. The verification results got much better after the reenrolment.

4.1.1 Parameters analysed

Accessed data to analyse was distance values, fingerprint matchtime, verification time and timestamp. Not all data were accessible at the current time. These are represented with a “null” character. Tables are presented in *Appendix F – Tables of data*.

The fingerprint matchtime is the time the algorithm takes to complete a decision in milliseconds. The analysis was to find dependencies between rejections and matching time. In the two following figures two users have made transactions and the matchtime are measured. The rejected verification attempts are marked in red.

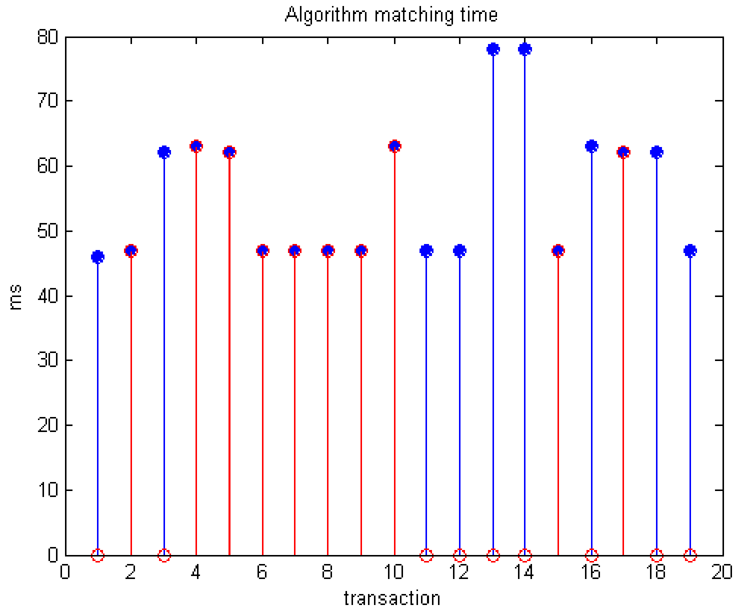


Figure 1: User$_1$’s matchtime is very even and does not have any influence on the rejections.

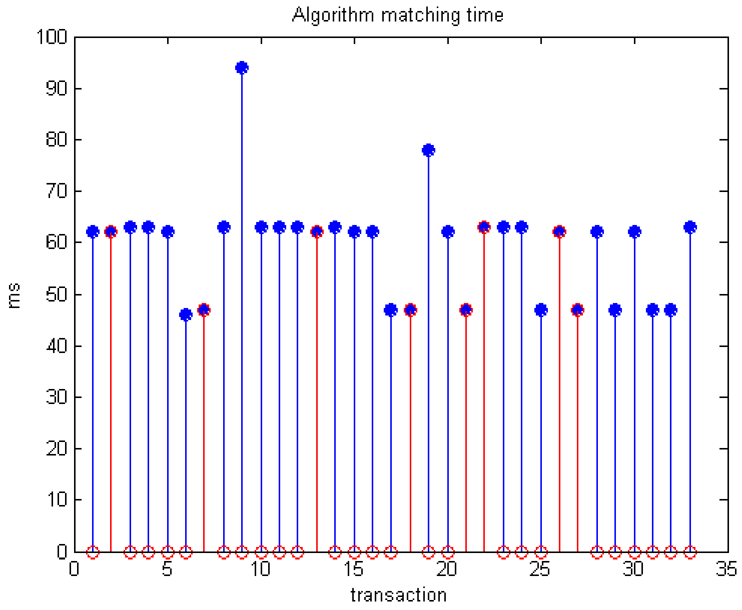


Figure 2: User$_2$’s matchtime is also very even and does not have any influence on the rejections either.

These analyses do not show any direct dependencies between rejections and the matchtime. The matchtime does not in this case supply any further information. An analysis to find dependencies between distance values and matchtime showed similar results.

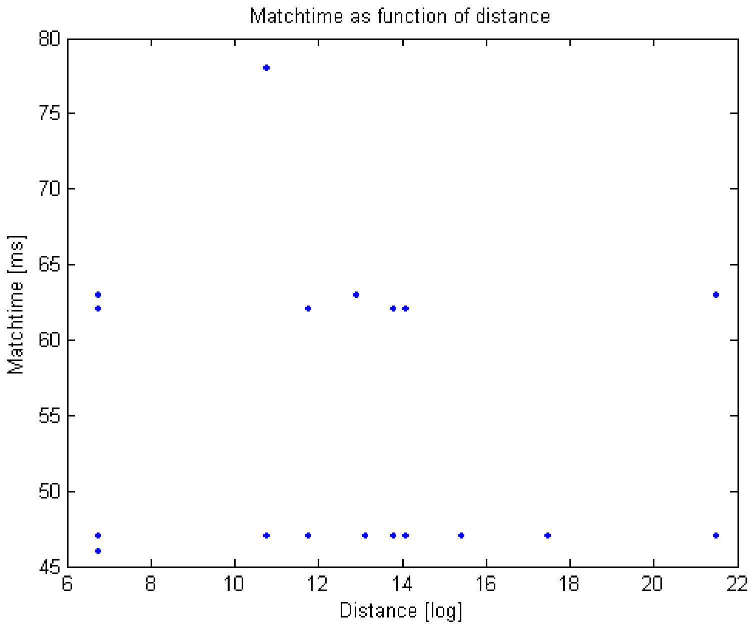


Figure 3:
User$_1$'s matchtime does not have any dependencies of distance. Each distance has many variations of matchtime.

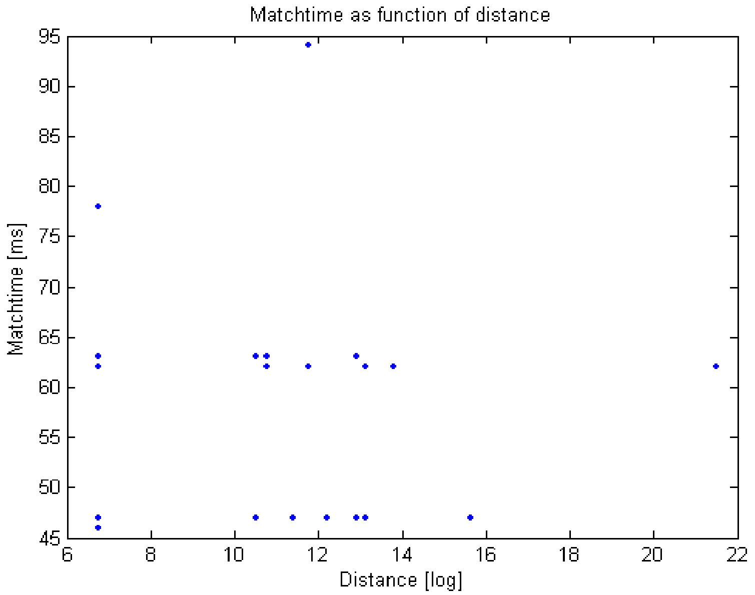


Figure 4:
User$_2$'s matchtime does not have any dependencies on distance either.

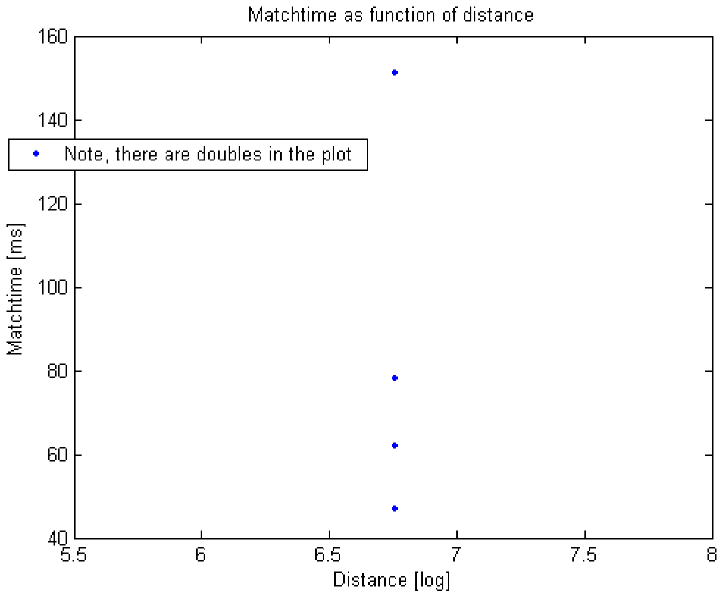


Figure 5: User3 have received only perfect distance values, but the matchtime is spread from 47 to 151 ms.

No dependencies appear when matchtime is compared to distance. By analysing biometric algorithm's matchtime, the conclusion is that there are no direct connections between matchtime and distance and the matchtime has not got anything to do with the rejections. Therefore the matchtime can not be used as a parameter.

The verification time is the time the whole procedure takes, from viewing the application on screen to the completed verification attempt. Unfortunately there were not many transactions recorded with belonging verification time. This parameter does not supply any trustworthy information. A user may have been distracted while verifying and therefore have the process taken longer time. The only remarkable thing was that the verification time was much shorter when a rejection occurred but this information is not important then searching for performance deviations. Verification time is illustrated below in figure 6.

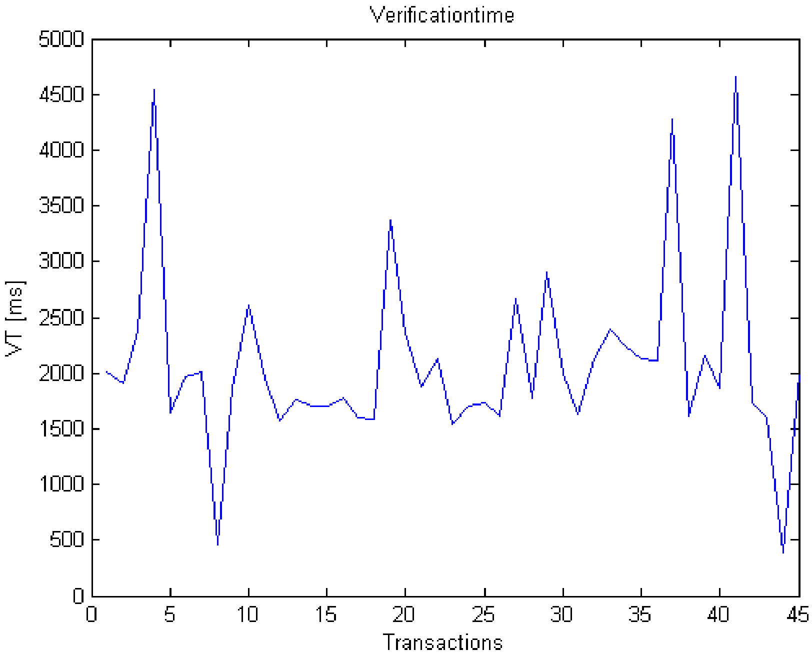


Figure 6: User4's verification-time (VT) does not seem to have any influences if the result is a rejection. A timeout sometimes occur but it has not happened here. The two 500 ms VT are rejected values.

According to these assumptions, the only remaining variables were distance and timestamp. Transactions one by one do not give much information, neither do averaging, variance or skewness measures on the whole dataset. The transactions during longer time are uncorrelated. The correlation is within a short timeframe, under a specific verification attempt. These events happen sporadically. Predicting this kind of data is impossible. A larger perspective has to be viewed being able to find abnormalities. The user statistics over time was calculated.

To start with, the period which generally fits most people is weeks. People usually have routines which are week-based. Therefore weekly calculations are done of user results. A calendar was implemented which simplified the process of weekly analysis.
A single histogram calculation with all values included will not be time dependent on the result. A weekly histogram of distance and timestamp was assembled and the mean values of each histogram were calculated. They are now time dependent and have better distributions represented. The distance distributions from the current fingerprint reader are not very similar to the model explained in Chapter 2 User distributions. The timestamp distribution is illustrated in *Appendix F – Tables of data*. Two users (same as before) mean weekly distance histograms are viewed below.

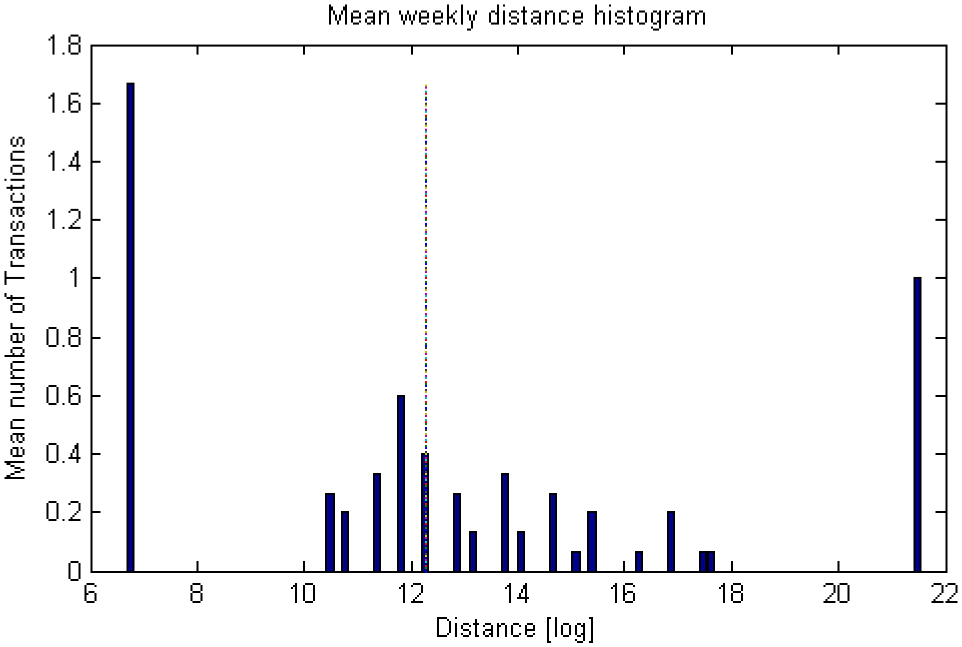


Figure 7: User$_1$ have not used the biometrics very much during the data collection. User$_1$ have been rejected very often and has very spread distribution. This is the goat explained above.

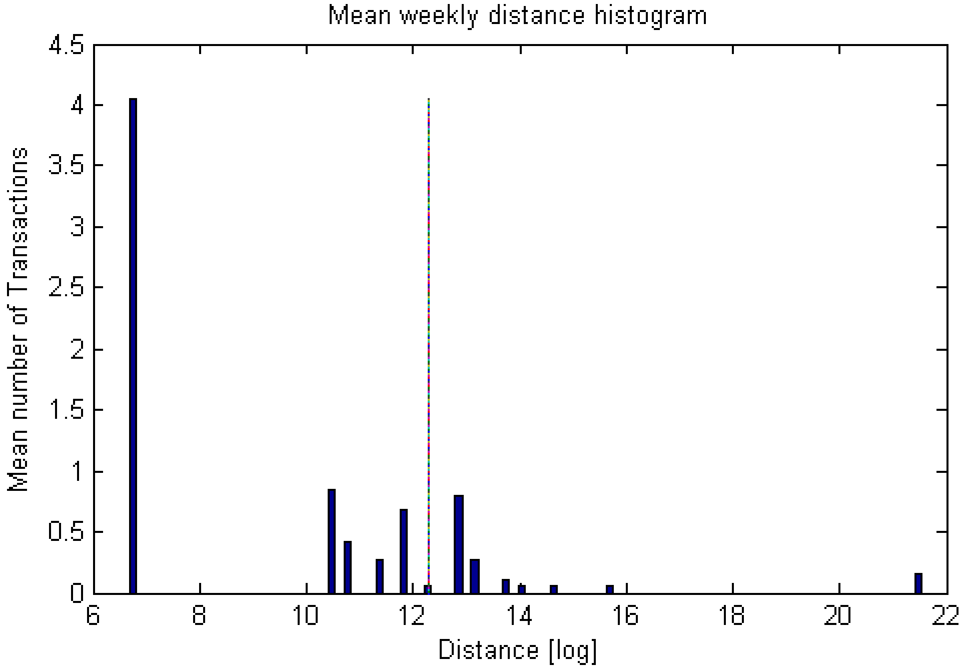


Figure 8: User$_2$ has got better results than User$_1$. User$_2$ have also used the biometrics more often than User$_1$.

The personal rejection rate is very interesting. This explains a lot about how well the biometric usage is going for that individual user. These measures are under supervision hence all results are caused by the volunteers themselves. The number of transactions is important to take in aspect to the rejectionrate, i.e. if the user has a rejectionrate of 30% and does in general three transactions a day then it is not a very big deal but if the user makes 20 transactions a day then the biometrics will become uncomfortable to use. It is important to include the number of transactions in the statistics to make a presentation of the personal rejection rates. Two personal rejectionrate statistics will be illustrated below. The first is a daily transactions counter with the corresponding number of rejections. The collection started week 38. It is weekly presented with each days result by itself. The thin bars are the number of transactions and the fat represents the number of rejected values each day. The point is to get an overall view of past activity in a graphical way. See figure 9 and 11.

The weekly amount of transactions with the rejection proportion is also demonstrated. It is basically the same thing as the daily verification result presentation but all days results are summed in current week. This will give a good trend curve representation of the rejectionrate. This is also illustrated separately. See figure 10 and 12.

User$_1$

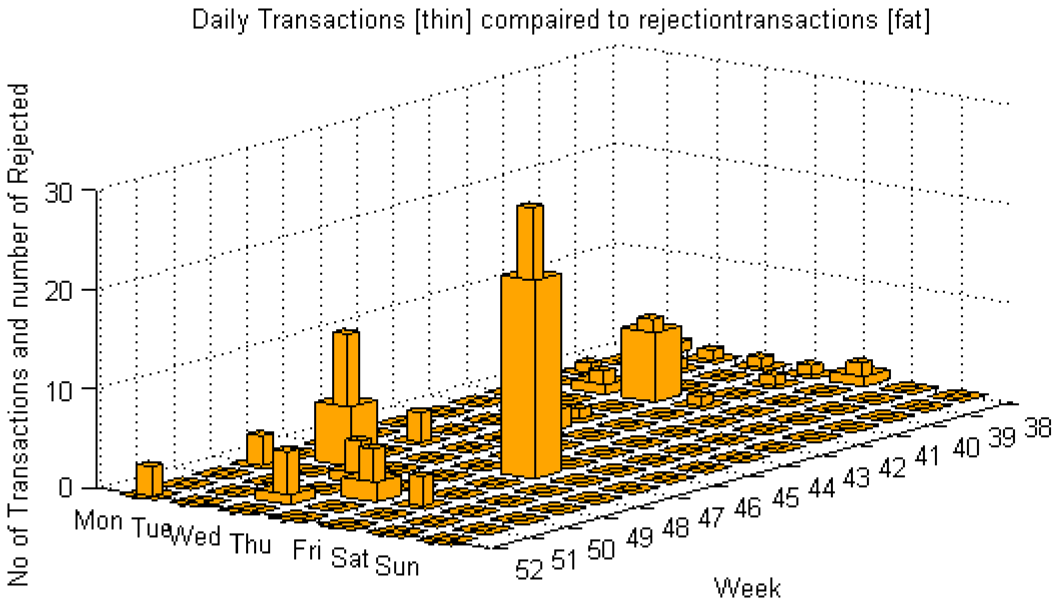


Figure 9: User$_1$ Daily transactions

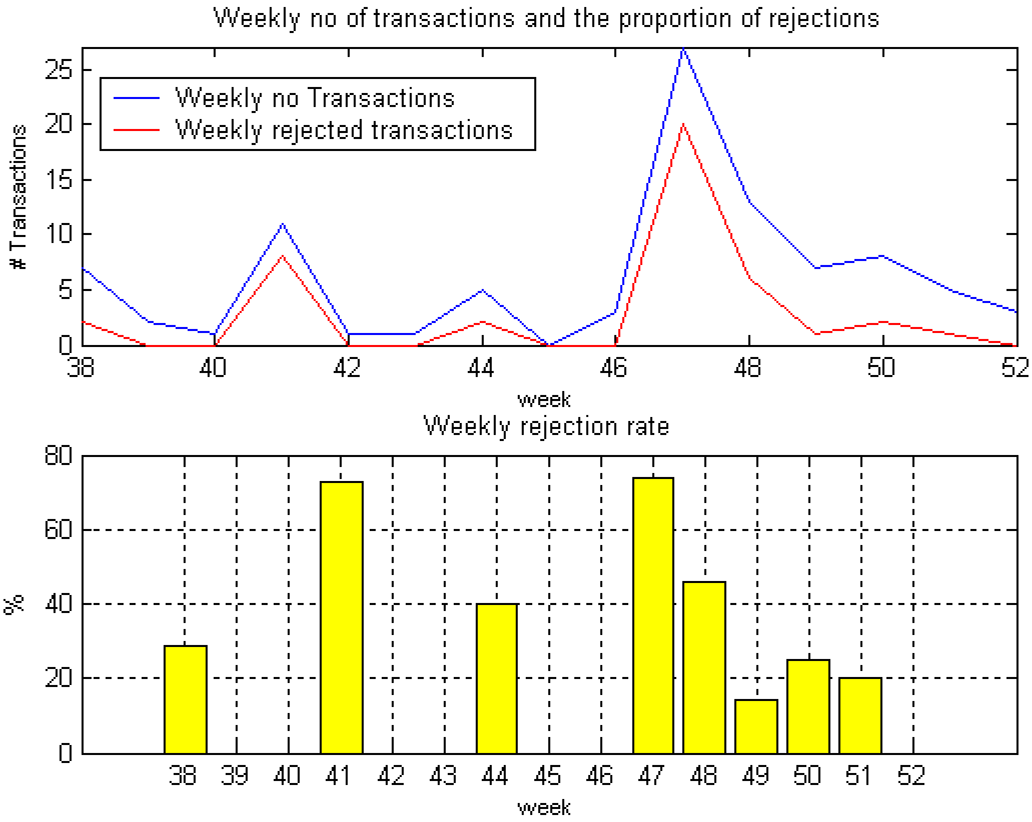


Figure 10: User$_1$ Weekly no. of transactions and proportion and Weekly rejection rate

This user reenrolled week 47 because of verification problems. The verification results were better afterwards but far from perfect. There is no answer to why this is happening to this user. The conclusion is as mentioned before that a “goat” is detected. This is regarded as an extreme case. See the big pike week 42 in Daily transaction figure. This occurred within a couple of minutes then the biometric refused to accept User$_1$.

User$_2$

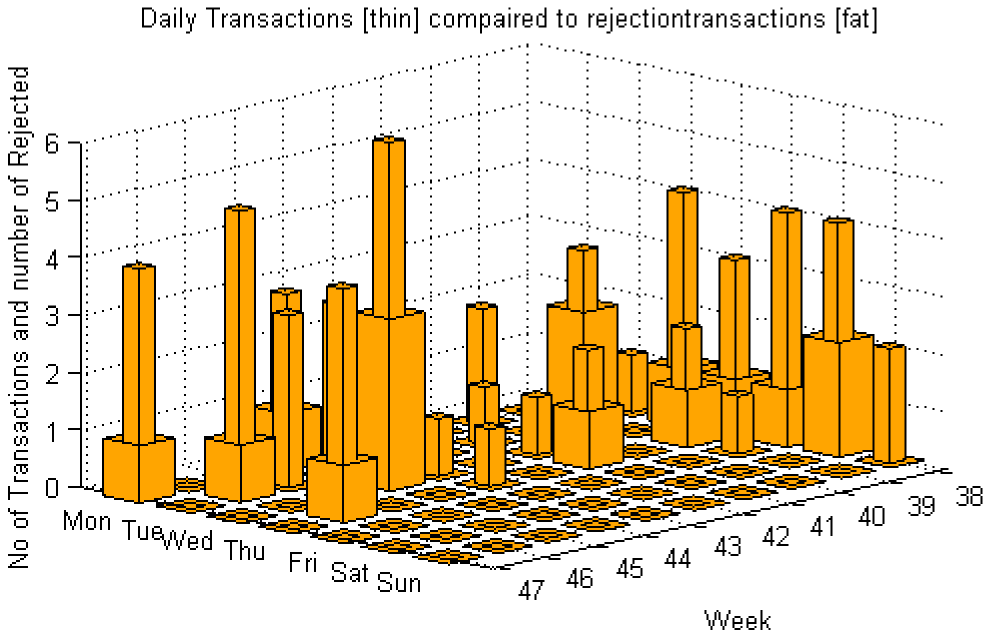


Figure 11: User$_2$ Daily transactions

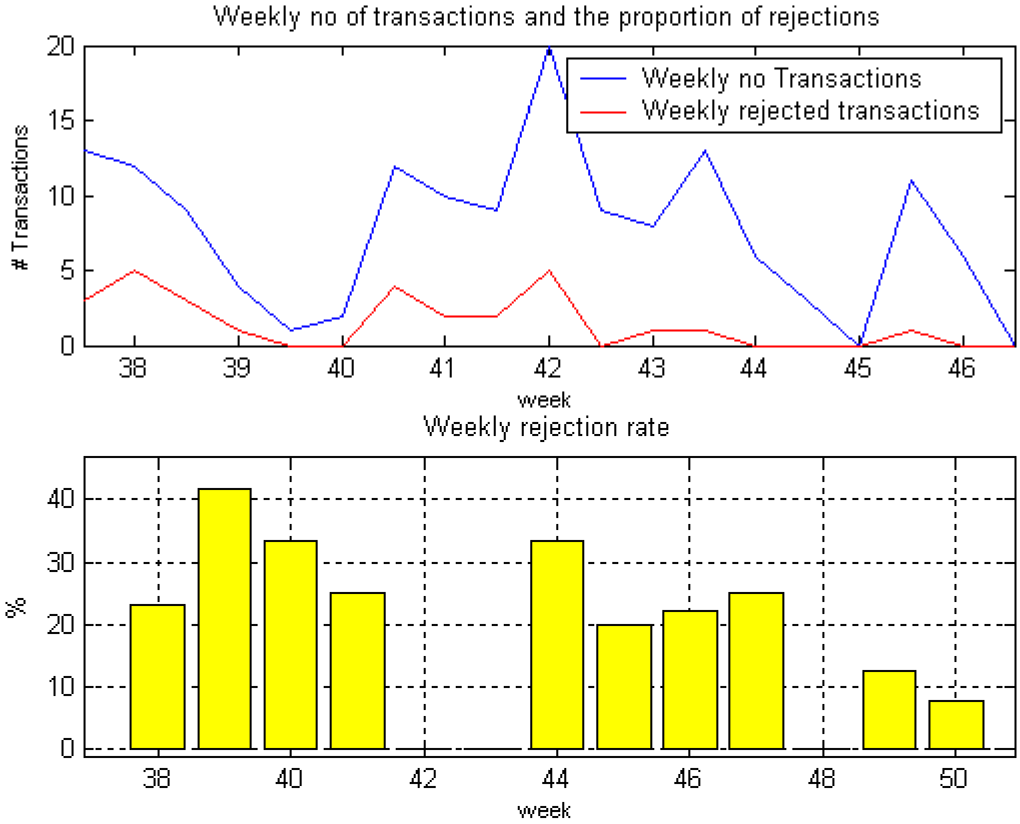


Figure 12: User$_1$ Weekly no. of transactions and proportion and Weekly rejection rate

User$_2$ have good results but is being rejected sometimes. The biometric average usage is approved. The rejectionrate is low when the user has done many transactions. User$_2$ can be seen as an average user.

4.1.2 Simulation

Application development would be delayed if an offline database had to be completed. OBL had prepared the project with a simulation of a long term personal usage programmed. It was developed to be compatible with the database. The user represents to use a biometric device as login on a computer. Analysis showed that score distribution was a bit unrealistic verification results. The simulation returned a fictive score with a range of 0 to 50 with a threshold set to 25. Hence values assembled from the biometric fingerprint reader were used to modify the simulation to be adapted to the real values. These were declared as distance values instead of score.

The simulation was used in the beginning especially in the long term prediction objective. The simulation was not that handy in abnormality detection case.

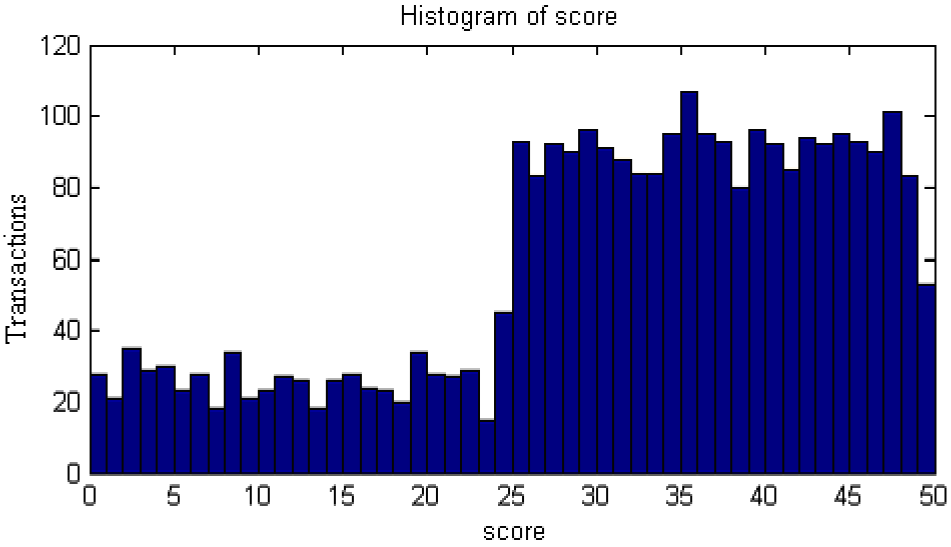


Figure 13: The original simulated usage during one year. This score distribution is formed into two uniformly distributed parts, the rejected part at the left and the accepted part at the right. This is not fairly reflecting the reality. The threshold was set to 25 and that is a very low security level. This distribution does not fit the model described in chapter 2. An application applied on this data would not be appropriate to use because usage of real systems does not have these distributions.

The offline data is stored as distance values. The distance values returned from the fingerprint reader have a wide distribution range which makes the values hard to survey. The best distance found is 859 and the worst 2147483646. Logarithmization is suitable as compression technique which simplified the overview. The distances seem to be codes but analysing these to figure out them has been left out. For translations between the linear and logarithm value see the table at *Appendix F – Tables of data*.

The resulting histogram of the modified simulation and PDF used is viewed on next page. This simulation is based on a Rayleigh distribution and the distances was organised in the order they usually appears. The distribution is provoked to change appearance described in Chapter 2 Template aging description.

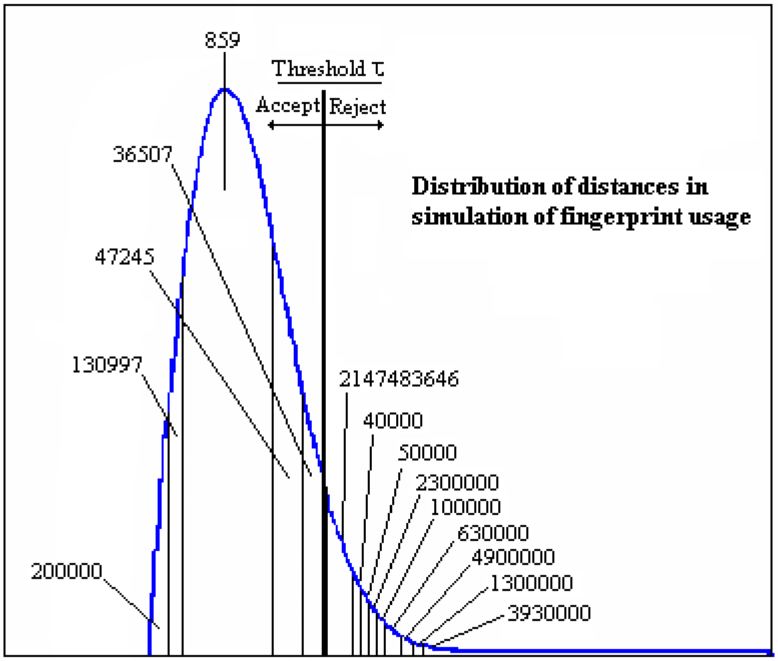


Figure 14: This is how the distributions were mapped to represent the real distribution in the simulation. A Rayleigh randomiser was used to determine the results.

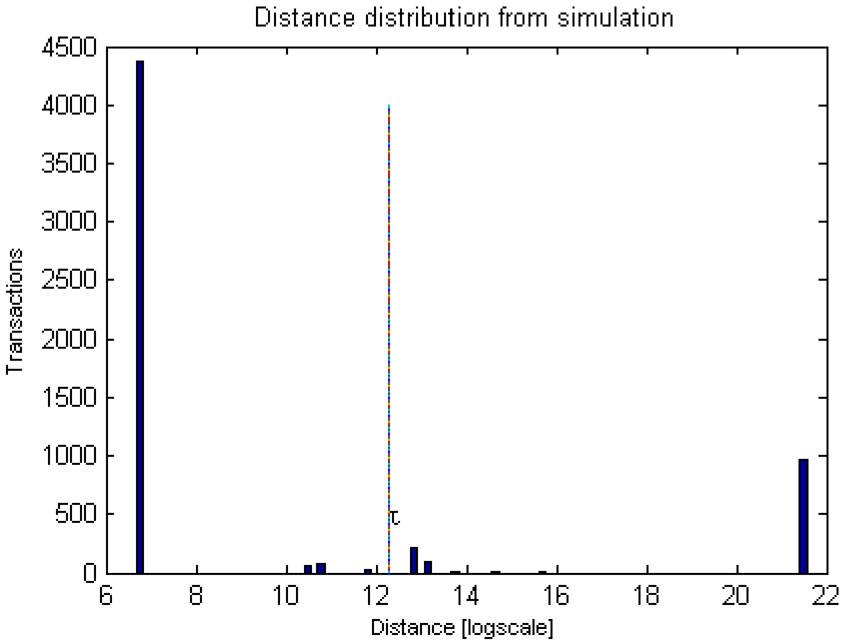


Figure 15: The result of the long term usage simulation. The distribution tends to be realistic in sense of an average user who sometimes will be refused access by the biometrics. Between the 900 to 1000 transactions has the rejectionrate becoming a bit higher because of the simulated provocation.

4.2 Unacceptable problem forecast using neural networks

The idea of forecasting possible problems became unthinkable after the data analysis was done. The data from transactions is far too uncorrelated and the degrees of freedom of usage are too many and the results are heavily dependent on them. There are very few parameters to work with being able to draw any conclusions ahead about what is going to happen. Neural networks are unlikely able to handle a scenario they have not been facing before [5]. This data analysis was done on real users but was collected during a very short period of time. This was not enough so a simulation was necessary beginning work with. Based on genuine user attempts and a theoretical template aging approximation data for a year time of usage were simulated. The transactions were as correlated as in a real life case. The backpropagation algorithm was tested to train the network. The results were unsuccessful but it was expected. The prediction of one sample ahead worked well of the values already collected. But prediction of multiple samples a head of desired data did not work at all. The need for a very long data sequence was also required to train it well.

In other fields where predictions have been done, like stock market predictions, there is data stored for a very long time. There are always relating factors affecting if the market goes up or down. If an economy crash happens in America a certain effect also happens in Sweden for instance. Both in the stock market and the gas load case, there have been continuously incoming data, which is not the case in this project. Transactions occur sporadic and there are as many combinations as there are users. This results in very high difficulties to model a general prediction procedure that fits all users.

After considerations of the results and the winning purpose of this project this task was cancelled for further efforts because there were no reasons to continuing. Time was running out and there are probably more optimisations left for the prediction program but no need wasting more time on it. Then faced reality this experiment would not have been given highest priority to begin with. This prediction of the rejectionrate would not as an application been very efficient or trustworthy.

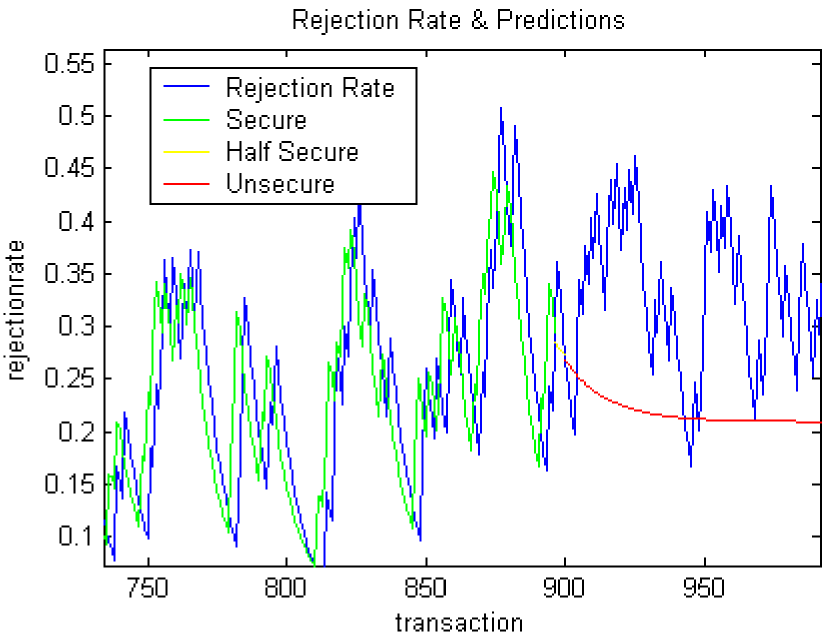


Figure 16: An illustration of the predicted stages of the user's rejectionrate (blue). The green line or so called secure is a prediction of the recorded data. The half secure (yellow line) shows where output samples are beginning to be input signals recursively. Unsecured line (red) shows where the forecasting begins without any knowledge from stored samples. Where the red line starts the validation set also starts. The result is not very impressive but that was expected.

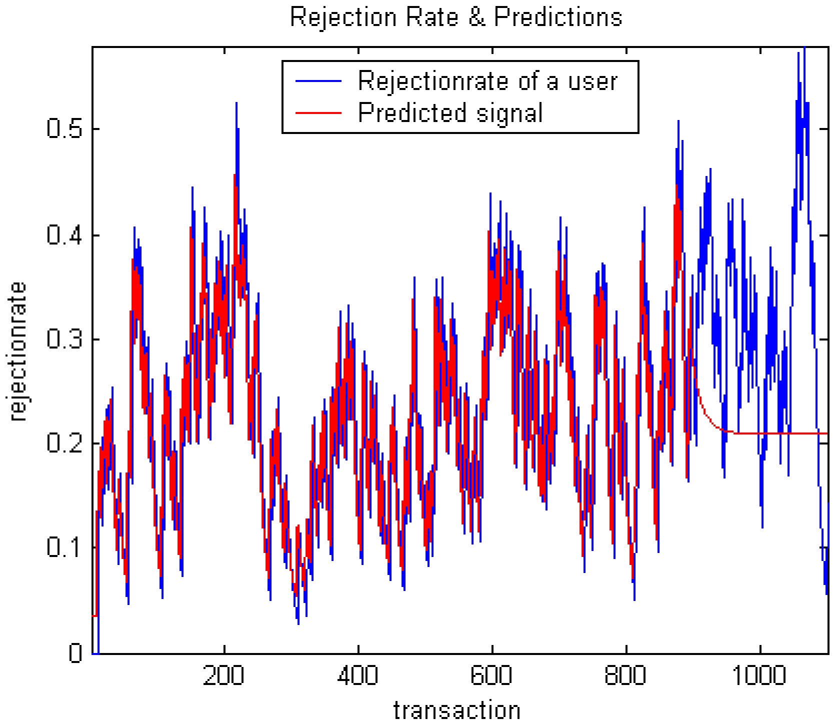


Figure 17: The whole signal. The last blue part is an illustration of the problem which is supposed to forecast and prevent.

4.3 Abnormal behaviour detection application

The abnormality detection application is the resulting layer on top of biometric level. The conceptual model was implemented and tested on offline data and composed scenarios. The application finally was implemented as a real time demo and tested in two smaller scenario evaluations. The first was not that successful and the program was calibrated and then tested a second time with satisfied results.

4.3.1 Offline testing

The application named Sheepdog is tested on simulations during development to achieve desired results. The tests were to performing different combinations of possible events which could occur in reality. The assumptions about biometric usage are made with specific care of the results from data analysis. Assumptions about other abnormal behaviour as the time aspect are made by common sense. To test this were fictive timestamps used to represent when the genuine user would not normally use the biometrics. Combinations of abnormal activities were tested with satisfied results. The results were relevant according to other users with different characteristics. The program returns a detailed report of all abnormal behaviour detected which has passed the output filter. If Sheepdog decides is that there are too many deviations from the normal behaviour a fraud possibility rate is included in the report. This is not always included because a low fraud possibility rate is not important information because the responsible tends to be the genuine user. This could confuse the administrator who is assumed not to be familiar of the program’s inner structure.

The results from a specific event can be various depending on the individual user’s pattern template and can vary from very high to low fraud possibility. The system’s rules adapts to every singe user. If not the entire template is known then it is difficult to forecast the results. The users were tested individually with different scenarios and the outcome was adjusted to be coherent with the results from the other users regarding the same event. It is not certain just because a user with very good score values and low rejection rate will get higher fraud attempt possibility than a user with opposite patterns. The results are very dependant on the incident's time. If it is lesser forms of abnormal events it is difficult to give an absolute answer to the event's outcome for each individual. Depending on the situation can the final decision be everything from nothing found to very high possibility of a fraud attempt.

Extreme events usually got similar results for the entire population. An example of an extreme event is an attempt with many rejections at eight a clock pm. The report contained information that the user has got verification problems and is very late occurrence (office hours are assumed in this case). It is also reported that the user never manage to pass. When the expert users were the testpersons the fraud possibility was similar to $user_1$ described before in this chapter. Sheepdog believes that it was another user with a possibility of 86% respectively 84%. $User_1$ was expected to sometimes get problems but neither was he expected to make transactions at the current time. Both results indicate that the possibility of an intrusion attempt was high.

4.3.2 Scenario evaluation

Sheepdog was demonstrated in a scenario evaluation in real-time. There were four tests performed. The first was to investigate how the program reacted on a single successful attempt. The purpose was to see if unnecessary messages were reported. The second test was to test a verification problem by the genuine user G. The third was a single impostor attempt against another user. This demonstrates impostors I who tries an attempt and then go away, avoiding get caught. The fourth test symbols an impostor who really wants to be accepted. The impostor makes five attempts against another user.

User	G single attempt	G verification problems	I single attempt	I five strikes
1	-	1	4, 8: 49%	1, 2, 4, 5, 8: 65%
2	6	1, 3, 8: 49%	3, 4, 6, 8: 49%	1, 2, 5, 8: 65%
3	-	1, 3, 8: 50%	3, 4, 8: 50%	1, 2, 5, 8: 65%
4	3, 8: 49%	1, 8: 49%	3, 6, 8: 49%	1, 2, 4, 5, 8: 85%
5	-	1	4	1, 2, 4, 5, 8: 65%

Scenario evaluation

	Codes
1	User was rejected more than usual
2	User may have verification problems
3	Abnormal distance value
4	User did not complete the verification
5	Too many rejections
6	Abnormal day
7	Abnormal time
8	Impostor attack possibility %

The results were not expected. The real-time application had never been tested on live transactions. It has only been tested on offline data and some runtime errors occurred. That resulted in some transactions were divided into two different attempts hence analysed separately. This is depending on database update problems. The results from analysis were a bit different from offline testing. The results depend much on the occurrence. In this case were the evaluation performed at a time when the users usually make transactions. For instance User 2 does not usually perform transactions on the current day hence an alarm was sent.

User 4 achieved an unwelcome message because of an unusual distance. It was concluded there was too much focus laid on distance. The data analysis on the vague datasets was misleading at this point. The theoretical assumptions were difficult to apply on this specific biometric system. The result User 4 achieved is not acceptable because usage will be very frustrating if these kinds of messages often appear. The distance is considerate to have less influence on the program because it would be uncomfortable to use the biometrics if possibilities of an impostor attack as high as 49% on a single successful attempt by a genuine user.

The verification problem test gave satisfied results according to the templates. User 4 and 5 has a bit higher rejection rates hence the program has considerate the results according to this characteristics.

The verification results were underestimated as parameter. This is shown especially in the results of the impostor attacks. The fraud decisions were too low. The number of rejected transactions should have higher influence of the decisions than before.

This evaluation was performed similar to a scenario evaluation with biometric devices to test different thresholds and find the equal error rate. The desired threshold is as high as possible with as low false rejection rate as possible. The same thing is desired with Sheepdog's algorithms. It must be correctly balanced so the alarms are justified and no false alarms occur. The scenario evaluations detected the software's weaknesses and where the algorithms needed improvements.

4.3.3 Final tests

After the scenario evaluation the program was corrected and adjusted and then a new test was performed. This time it was acceptable results. The results can be viewed below.

User	G single attempt	G verification problems	I single attempt	I five strikes
1	-	1	4, 8: 80%	1,2,4,5 8: 80%
2	-	1	4, 8: 80%	1,4,5, 8: 80%
3	6,7, 8:36%	1,6,7, 8: 39%	4,6,7, 8: 85%	1,2,4,5,6,7, 8:85%
4	-	-	4, 8: 80%	1,2,4,5 8: 80%
5	-	1	4, 8: 80%	1,2,4,5 8: 80%

Second evaluation

The results are relevant according to the users' templates. The program is indeed not perfect and needs optimisations but the initial problems were corrected. The distance analysis was left out for future versions. All impostor attempts got high rates. If the time performing the evaluation not had been performed at office time which is ordinary for these users, it should have been even higher. A remarkable detail was User 1-3 should have even higher impostor possibility because of specific information in their templates. User 3 is not a very active user of the biometrics and at the testing was abnormal time for him to verify himself. That explains the higher impostor possibility.

Sheepdog is more forgiving against users who have problems but succeed to verify at last. A user tested to verify on a rare time. He was rejected three times before manage to verify. This looks very suspicious to Sheepdog which calculated it to be 49% possible intrusion attempt.

The major parts of available data recorded in database are from transactions made by professionals within biometrics. Transactions made by them are mainly done in experimental purposes hence the real time application was harder to test. Even if the data is not manipulated there are differences between test the application in a lab and validate results from an operational system. Because of this situation the conclusion was to get it running and test the responses from known events stored in database. Further validation of Sheepdog can be done when Optimum preCon^{TM} is installed into an operational system with live users in their natural environment.

Sheepdog performs detection of template aging similarities at the same time as estimating templates. Here was just enough data to estimate a template hence to test aging effects then simulations were necessary. The simulations are based on every user's template which forms individual long term simulations of usage to test the application against. Results of simulations were Sheepdog is able to detect changes of distributions of distance values are found in *Appendix E - Radical distribution change detection*. This is also tested on real values. The problem was as mentioned before, the lack of assembled data from an operational installation. In the real time application it is much more important that the data is collected from natural usage, which is not the case here. Therefore is the testing on the simulation as accurate as the test on the collected data. The test on the collected data showed similar results as in the simulation but some smaller parameter adjustments were necessary.

The absence of previously recorded live data from an operational biometric system complicated the project. The algorithms used were mainly developed based on simulations which were designed based on observations during data collection. This made it difficult to test the application, especially in sense of long term surveillance. The simulation tests are established as the ground for further development.

Chapter 5

Conclusion

This thesis work has started a development of an application with an innovative thinking. The project's final program, Sheepdog, acts as an additional intelligent layer above the biometrics for increased security. It is working independently of the chose of biometric and solves inherent problems that biometrics bring. In the biometric level is it impossible to confirm that it was the genuine user, who was hard to recognize and therefore rejected; now a possibility developed in Sheepdog called Abnormality Detection. The Abnormality Detection model became the fundamental part of this thesis work. The theory and experiences worked as basis for investigation of biometric usage with abnormal activity. It resulted in a program which statistically tries to recognize the user when they use the biometrics with the assumption that biometrics are not perfect and may causing false rejections. The degree of freedoms in sense of usage is many. Therefore the decisions made by the application are always balanced between the doubt of an impostor attempt or a genuine user with verification problems and when the program decides that it is very likely that it was an impostorattempt an alarm is sent, otherwise warnings. There is also an algorithm implemented to find if a score or distance distribution is changing and in which direction. It worked really well on simulations of template aging like effects on the user.

All ideas from the thesis proposal were not equally successful and therefore further research was deferred to later projects. This was the long term prediction of the personal rejection rate being able to discover template aging effects long time ahead. A template aging experience is first needed and this can take long time. Neural networks was considered the best prediction tool to use because ability of learning a curve. They worked well for short predictions but unfortunately not equally well for long forecasts ahead as needed. This experiment was rewarding in many aspects anyway. When a more sustainable basis exists, the preparations are already made.

Optimum preCon^{TM} together with Sheepdog is a powerful complement to invested biometric systems. Confirmation of likely impostor attempts is very valuable information especially when the security level is high and knowledge about the system's usage is desired. Information about the biometric systems efficiency is enhanced. False alarms are avoided and maintenance of the biometric usage is now available simultaneously.

5.1 Future works

This is a new concept hence further research is required. The algorithms need optimisation or adjustments for improvements.

This layer is supposed to be working against any biometric systems. So far has only a single fingerprint reader Precise 100 MC from Precise Biometrics been used for data analyses and the application is only tested on this device.

The software is tested against simulations but not in operational mode on Optimum preCon^{TM}. There were not time enough for a single thesis worker to complete full functional software adapted to the server program. A demo version exists which is mentioned to demonstrate the results. A program triggered from the database is suggested instead of polling occurred events from a client. Optimum preCon^{TM} and Sheepdog should work together. Today the alarms caused by Sheepdog are not implemented to be viewed in Optimum preCon^{TM}. The reported messages are very analytic and these details may possibly be minimized in amount. This is suggested to be an advanced setting in Optimum preCon^{TM}.

After the second evaluation test the algorithm achieved desired results. A larger population to test Sheepdog is required and the data must be recorded live from an operational biometric system achieving better confidence intervals that the program really works. There were not any possibilities to calculate any confidence intervals during this project. The small number of users to test the algorithms was simply too few.

Design a database for template storage is more efficient than a file based storage. Further compression and encryption of the template is recommended.

Glossary

BioAPI	Standard for biometric application developments
Biometric	Analyse of biological data, often to identify a person
Enrol	Registering in a biometric device
False acceptance rate (FAR)	Percentage of impostors incorrectly matched to a valid user's biometric.
False rejection rate (FRR)	Percentage of incorrectly rejected valid users.
Histogram	A diagram were each bin on x axis is a specific value and the y axis the frequency of each value
Identification	The system compares the user against others in database. One-to-many comparison
Neural Network	A mathematical model, often used in recognition purpose. Have ability to learn by a specified algorithm
Prediction	Methods to forecast future values or trends.
Probability Density Function	Shortly named PDF. Describes which distribution of values a variable can take. Gaussian distribution is a common example.
Retina	The inner membrane of the eye globe
Template	A mathematical representation of biometric data stored to a database by the biometric device when enrolled
Transaction	A login on the biometric device stored in database
Verification	User trying to log in a system claiming being him/her-self. One-to-one comparison.

Bibliography

[1] A.J Mansfield & J.L Wayman, Best Practices in Testing and Reporting Performance of Biometric Devices, 2002

[2] James L. Wayman, Error Rate Equations for the General Biometric System, US National Biometric Test Centre, March 2000

[3] P. J Phillips, P. Grother, R. J. Michaels, D. M. Blackburn, E. Tabassi, M Bone, Face Recognition Vendor Test 2002, Evaluation Report 2002

[4] U. Uludag, A. Ross, A. Jain, Biometric template selection and update: a case study in fingerprints, 2003

[5] C. Lee Giles, Steve Lawrence, A. C. Tsoi, Noisy Time Series Prediction using a Recurrent Neural Network and Grammatical Inference, 2001

[6] Michael E. Schuckers, Some Statistical Aspects of Biometric Identification Device Performance, Stats Magazine September 5, 2001

[7] David M. Skapura, Building Neural Networks, Addison-Wesley [ISBN 0-201-53921-7]

[8] Philip D. Wasserman, Advanced Methods in Neural Computing, [ISBN 0-442-00461-3]

[9] Per Cornelius, Neural network course compendium ETD 007, Blekinge institute of technology, Sweden

[10] G. Doddington, Sheeps, goats, lambs and wolves: An analysis of individual differences in speaker recognition performance, ICSLP'98, Sidney, Australia November 1998

[11] M. Sandström, Liveness Detection in Fingerprint Recognition Systems, LITH-ISY-EX-3557-2004, http://www.ep.liu.se/exjobb/isy/2004/3557/exjobb.pdf

[12] N.H Viet, J.Mandziuk, Neural and fuzzy neural networks for natural gas consumption prediction, pp 759-768, 2003

Figure references

B. Goudarzi Pour and M. Zackrisson, Facial Recognition Biometrics, applying new concepts on performance improvement and quality assessment, May 2003

Appendix A - Template estimation

A.1 System Identification

When Optimum preCon™ is installed it cannot solve any intelligent tasks immediately. It has to be running for some time and collect data from users before the algorithms are able to make any conclusions. A system identification is important to do on each individual using the system otherwise is it impossible to tell if something is an abnormality for the specific user. The system identification task is to form a pattern model for each user. These patterns define what a normal behaviour is and that everything else is not suitable behaviour and that can be classed as an abnormality. It is called a pattern **template** or shortly **T**. All abnormalities are represented in points and decisions are based on thresholds for each parameter. The thresholds control the security levels. The points are set on negative incidents and generate points at each suspicious moment. A single abnormality may not be as unusual to be classified as a critical case, but a few pattern breaking parameters together can raise this to a critical level and call red alert. This is based on a verification system, i.e. a 1:1 comparison.

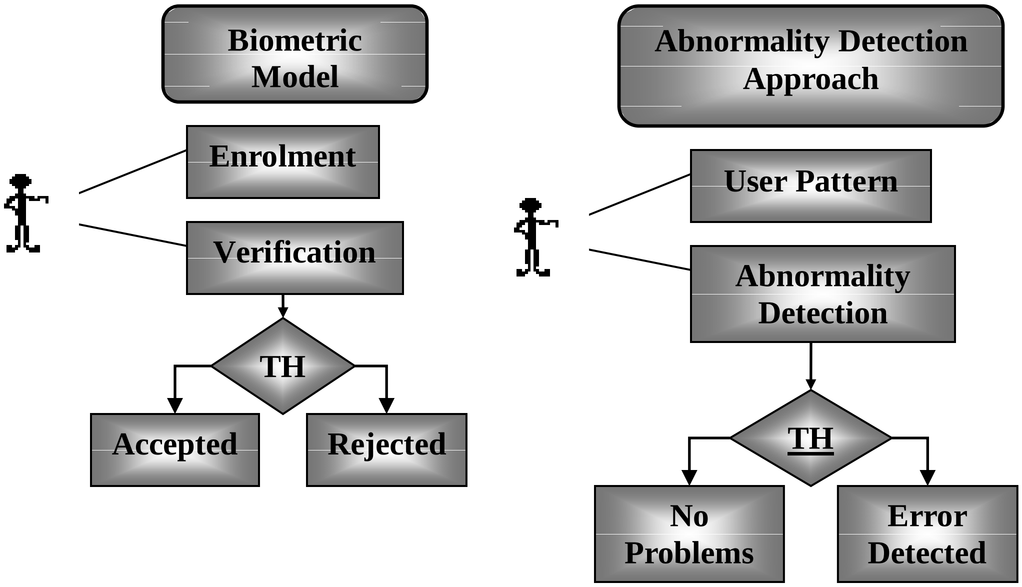


Figure 1: A symbolic abnormality detection procedure is based on the biometric verification procedure model. Thresholds (underlined) in this approach are multiple. This forms a kind of multimodal behavioural biometric.

An overview of the system identification is illustrated in the figure 2 below.

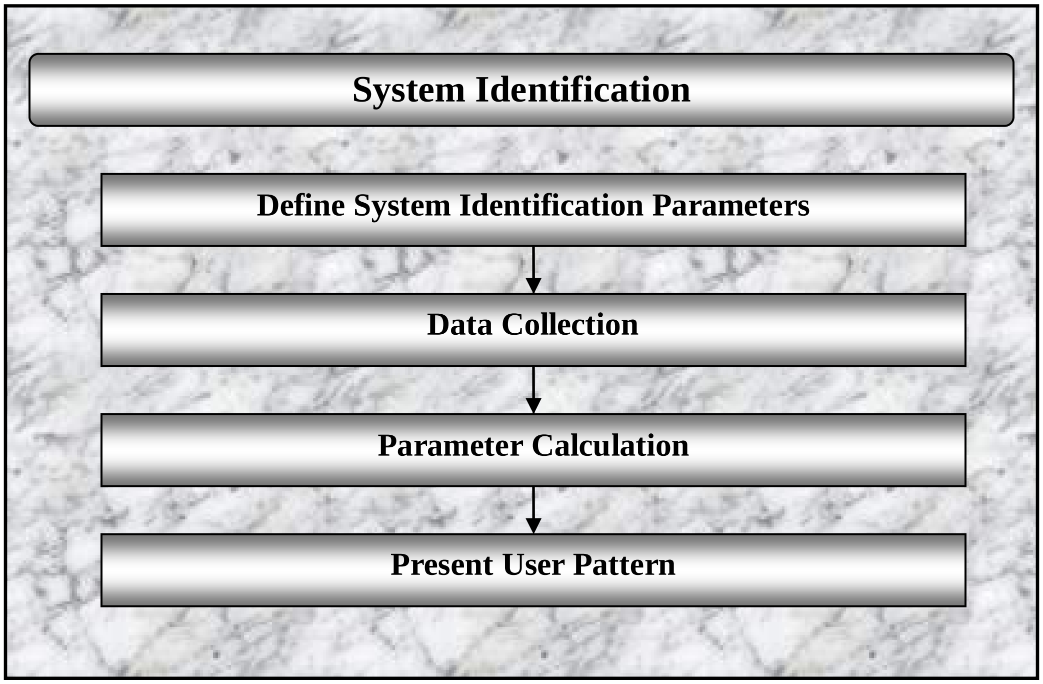


Figure 2: The pattern template is created through the steps showed in the figure.

Appendix B - Application structure

This application can be described as a multimodal behaviour biometric. The application's goal is to identify abnormal behaviour on individuals using the biometric system. The status of a user is taken into account the fraud attempt possibility decision. All users have their own unique userpattern. The rules are adapted to every single user. The template described in *Appendix A - Template estimation* is used to determine if an event fits into an ordinary behaviour.

B.1 Intelligent system presentation

The application's work name is **Sheepdog**. A name inspired by a CPU's watchdog and the purpose of the application, guarding sheep, lambs and goats from wolves (see *Chapter 2 User characteristics*). The application also discovers if a person has problems using the biometrics.

B.2 A summary of performance

Sheepdog is able to detect:

- Radical usage which seems to be trespassing attempts
- Users who has verification problems
- Changes in usage over time

Appendix C - Abnormality detection modules

This appendix will in later versions describe the functionality of the AD (Abnormality Detection) module in detail. This is because of confidential matters. This module always is individually adapting to the templates **T**. This module contains algorithms to detect abnormal activity among biometric users.

C.2.1 Future versions

Some suggestions of further research and upgrades are

- Liveness detection information
- Threshold settings
- Integrated schedule
- Station sensitivity

Manufacturers develop better biometric devises which can perform liveness detection. Some biometrics are fooled by artificial or cadaver limbs. An example of this is that some face recognition devices are fooled by photography of an authenticated user is held in front of it. Gelatine fingerprints are a popular topic at the moment [C.1]. Information from liveness detection algorithms is very valuable in purpose of investigating the authentic usage.

The threshold level may contribute in decisions. If knowledge about high or low level of the threshold, some decisions is simplified. False acceptance is not equally likely when threshold is high compared to a system with lower thresholds and lower security.

An integrated schedule is desirable. Knowledge of when a user should be away from office is valuable if events occur anyway. That would seem very suspicious.

Station sensitivity is a suggestion of future work where some information regarding the placements of the biometrics is involved. An initial concept is if the biometrics is far from each other then is it impossible making transactions during a short time.

C.2.2 Bibliography

[C.1] M. Sandström, Liveness Detection in Fingerprint Recognition Systems, LITH-ISY-EX-3557-2004, http://www.ep.liu.se/exjobb/isy/2004/3557/exjobb.pdf

Appendix D - Radical distribution change detection

This appendix describes the functionality of the DOC (degree of change) module, but the algorithms are not viewable in this version. This module performs an overview investigation of users' results. In exception to AD module is DOC a longer period analysing tool which investigate the overall usage. The motive is finding template aging effects. This is indeed not a template aging detector. That would be absurd to claim because it is an unsupervised system and template aging effects are estimated through strictly controlled studies, were all other defects are isolated. What will seem as aging effects can be depending on different factors like hardware defects, environmental factors and impostor attacks. But the goal is to detect these similarities because this is bad things anyhow.

D.1 Functional requirements

DOC must have information in form of score or distance values from transactions. Otherwise this module would be useless.

D.2 Activation

Optimum preCon^{TM} bootstraps Sheepdog in the late hours of Sundays. Then Sheepdog performs the diagnostics. It iterates through the user id database and updates the *degree of change flag* in template **T**. Sundays is the best suggestion of running the application because it captures the whole week activities, but 7-9 day periods would work fine too.

D.3 Degree of change flag

As earlier mentioned, biometric users usability to use biometrics may change during time. Important to remember is that all changes are not bad things. The user may have improved the results and then is the template reestimated because the older one is not suitable anymore. Sheepdog is not supposed to fall into the same trap as it is trying to protect the users' from[1].

Therefore is a parameter **degree of change flag** introduced. The flag contains information about user's status. It is defined as

Case 0: Nothing has radically changed
Case 1: Update template. User's ability to use biometrics has improved.
Case 2: Alarm! The user's distribution is not suitable anymore. User should reenrol soon.
Case 3: Red Alert! User has huge problems and must reenrol immediately!

[1]There is biometrics with template update. The issue remains if it is the real user that the device updates the template to. That chance is unlikely here. In those cases have an impostor succeeded every attempt on the user's template and received better results than the genuine user.

Appendix E - Transaction Identification, fraud possibility decision making

This appendix describes the solution of how the TI (Transaction Identification) module works. TI module makes a decision of the possibilities that the event was not caused by the genuine user based on the other modules results. This appendix contains algorithms, which are restricted to unauthorised readers.

Appendix F - Tables of data

This appendix contains data assembled from the database. The two most interesting users are presented in the tables. They are referred to as User$_1$ and User$_2$. Further on is a short table representing the most common user type; User$_3$ who seem to manage well. Finally is User$_4$ who has used a different application where verification time is accessible.

F.1 Tables

Table of distances and threshold

Linear	Log
859	6.7558
36507	10.5053
47245	10.7631
90194	11.4097
130997	11.7829
200000	12.2061
400000	12.8992
500000	13.1224
1000000	13.8155
2300000	14.6484
4900000	15.4047
6300000	15.6561
39300000	17.4867
2147483646	21.4876

Threshold	Log
214748	12.2772

Factors causing rejections
Wet fingers
Dry fingers
Dirty fingers
User fumble
Template aging
Intrusion attempts
Timeout
HW/SW
Environmental factors

User$_1$

Distance	VR	Timestamp	Matchtime [ms]
2147483646	1	2004-09-13 15:15:35	NULL
200000	0	2004-09-13 15:15:47	NULL
36507	0	2004-09-14 10:37:52	NULL
859	0	2004-09-15 09:28:33	NULL
36507	0	2004-09-16 09:33:34	NULL
2300000	1	2004-09-17 10:09:37	NULL
859	0	2004-09-17 10:09:46	NULL
859	0	2004-09-21 13:31:58	NULL
859	0	2004-09-23 09:54:19	NULL
859	0	2004-09-27 09:41:10	NULL
859	0	2004-09-30 17:33:21	46
4900000	1	2004-10-05 17:00:32	47
859	0	2004-10-05 17:00:41	62
2147483646	1	2004-10-06 15:45:44	63
1300000	1	2004-10-06 15:45:53	62
500000	1	2004-10-06 15:46:06	47
1000000	1	2004-10-06 15:46:28	47
1300000	1	2004-10-06 15:46:36	47
39300000	1	2004-10-06 15:47:07	47
400000	1	2004-10-06 15:47:25	63
130997	0	2004-10-06 15:47:34	47

47245	0	2004-10-07 10:32:22	47
47245	0	2004-10-11 13:41:34	78
47245	0	2004-10-20 15:59:17	78
2147483646	1	2004-10-27 13:30:56	47
859	0	2004-10-27 13:31:26	63
1000000	1	2004-10-27 13:31:42	62
130997	0	2004-10-27 13:32:03	62
859	0	2004-10-27 13:34:14	47
200000	0	2004-11-09 10:43:29	62
90194	0	2004-11-09 10:43:45	47
859	0	2004-11-09 10:43:55	47

User$_2$

Distance	VR	Timestamp	Matchtime [ms]
859	0	2004-09-13 11:51:31	NULL
859	0	2004-09-14 10:37:16	NULL
859	0	2004-09-15 09:17:08	NULL
859	0	2004-09-15 09:20:27	NULL
400000	1	2004-09-15 14:36:04	NULL
859	0	2004-09-15 14:36:16	NULL
400000	1	2004-09-16 09:06:14	NULL
90194	0	2004-09-16 09:06:24	NULL
90194	0	2004-09-16 16:46:44	NULL
47245	0	2004-09-17 10:08:39	NULL
2300000	1	2004-09-17 16:55:31	NULL
859	0	2004-09-17 16:55:39	NULL
36507	0	2004-09-17 16:56:55	NULL
2147483646	1	2004-09-18 11:51:16	NULL
500000	1	2004-09-18 11:51:33	NULL
130997	0	2004-09-18 11:51:42	NULL
859	0	2004-09-18 14:58:20	NULL
859	0	2004-09-19 12:17:07	NULL
859	0	2004-09-19 14:42:02	NULL
2147483646	1	2004-09-21 13:15:30	NULL
400000	1	2004-09-21 13:15:46	NULL
36507	0	2004-09-21 13:15:55	NULL
400000	1	2004-09-23 09:52:13	NULL
859	0	2004-09-23 09:52:24	NULL
859	0	2004-09-24 15:06:41	NULL
47245	0	2004-09-27 09:36:09	NULL
47245	0	2004-09-27 09:40:34	NULL
36507	0	2004-09-30 11:11:34	NULL
859	0	2004-09-30 11:12:53	NULL
859	0	2004-09-30 17:28:41	62
1000000	1	2004-09-30 17:28:53	62
859	0	2004-09-30 17:29:03	63
859	0	2004-09-30 17:35:58	63
130997	0	2004-10-05 16:59:53	62
859	0	2004-10-06 15:45:35	46
400000	1	2004-10-07 10:31:40	47
859	0	2004-10-07 10:31:50	63

130997	0	2004-10-11 13:17:17	94
859	0	2004-10-20 15:59:58	63
859	0	2004-10-21 14:43:49	63
47245	0	2004-10-25 12:38:08	63
2147483646	1	2004-10-25 16:03:09	62
36507	0	2004-10-25 16:03:19	63
859	0	2004-10-26 15:02:42	62
130997	0	2004-10-26 15:02:50	62
90194	0	2004-10-26 15:02:58	47
500000	1	2004-10-27 13:28:22	47
859	0	2004-10-27 13:28:30	78
859	0	2004-10-27 13:28:38	62
400000	1	2004-10-27 16:35:21	47
400000	1	2004-10-27 16:35:29	63
36507	0	2004-10-27 16:36:01	63
47245	0	2004-11-01 11:00:53	63
90194	0	2004-11-01 11:01:03	47
500000	1	2004-11-01 11:01:14	62
6300000	1	2004-11-01 11:01:25	47
130997	0	2004-11-01 11:01:32	62
36507	0	2004-11-01 13:42:44	47
47245	0	2004-11-01 13:42:52	62
200000	0	2004-11-02 10:51:19	47
859	0	2004-11-02 10:51:28	47
47245	0	2004-11-02 10:51:37	63
1000000	1	2004-11-09 10:41:55	78
47245	0	2004-11-09 10:42:03	47
130997	0	2004-11-09 10:42:10	63
859	0	2004-11-09 10:42:18	47
47245	0	2004-11-09 10:42:25	63
859	0	2004-11-11 16:26:02	47
859	0	2004-11-11 16:26:12	62
400000	1	2004-11-11 16:26:20	63
859	0	2004-11-11 16:26:28	62

User$_3$

Distance	VR	Timestamp	Matchtime [ms]
859	0	2004-09-30 17:33	47
859	0	2004-10-05 17:00	62
859	0	2004-10-06 15:47	47
859	0	2004-10-11 13:42	78
859	0	2004-11-02 14:33	151

User$_4$

Timestamp	Verification time [ms]	Distance	VR
2004-09-30 17:24:54	2016	859	0
2004-10-01 09:11:27	1905	859	0
2004-10-01 10:04:39	2372	859	0
2004-10-01 11:12:14	4546	859	0
2004-10-01 11:23:03	1641	859	0
2004-10-01 13:35:54	1966	859	0
2004-10-01 13:49:18	2015	859	0
2004-10-01 16:25:57	454	2147483647	1
2004-10-01 16:26:05	1859	859	0
2004-10-05 11:20:11	2610	859	0
2004-10-05 11:26:19	1954	859	0
2004-10-05 11:27:31	1576	859	0
2004-10-05 11:28:45	1764	859	0
2004-10-05 11:30:25	1702	859	0
2004-10-05 11:33:34	1702	859	0
2004-10-05 11:37:57	1780	859	0
2004-10-05 13:45:27	1598	859	0
2004-10-05 13:45:42	1583	859	0
2004-10-05 13:56:18	3373	859	0
2004-10-05 13:56:32	2359	859	0
2004-10-05 13:58:02	1875	859	0
2004-10-05 14:10:56	2133	859	0
2004-10-05 14:11:24	1538	859	0
2004-10-05 14:12:30	1710	859	0
2004-10-05 14:12:39	1726	859	0
2004-10-05 14:15:51	1614	859	0
2004-10-05 14:16:02	2664	859	0
2004-10-05 14:17:16	1771	859	0
2004-10-05 14:21:20	2899	859	0
2004-10-05 14:21:28	1991	859	0
2004-10-05 14:22:52	1629	859	0
2004-10-05 14:35:49	2131	859	0
2004-10-05 14:39:18	2396	859	0
2004-10-05 14:43:27	2240	859	0
2004-10-05 14:45:06	2130	859	0
2004-10-05 14:49:52	2114	859	0
2004-10-05 14:50:08	4277	859	0
2004-10-05 14:51:50	1614	859	0
2004-10-05 14:54:55	2162	859	0
2004-10-05 14:59:17	1864	859	0
2004-10-05 15:00:48	4652	859	0
2004-10-05 15:08:19	1739	859	0
2004-10-05 15:08:29	1598	859	0
2004-10-06 15:44:30	391	2147483647	1
2004-10-07 10:31:15	2001	859	0

F.2 Timestamp histogram

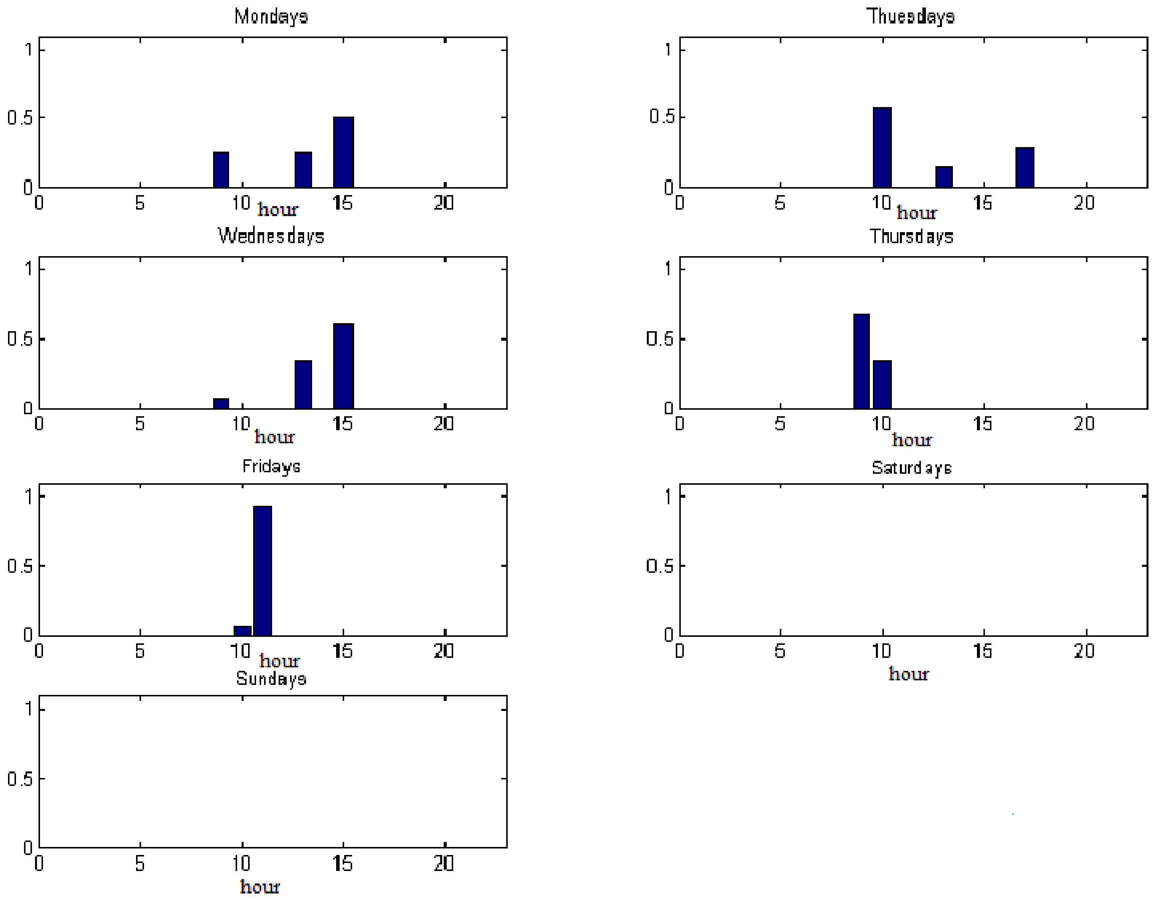


Figure 1: These are normalised histograms which form a weekly averaged usage of the biometrics. It is divided into the different weekdays.