The statistical strength of forensic identification through mobile phone call data records

Helena B.R. van Eijk

Master thesis Mathematics (public version)
Specialisation: Statistical Science for
Life and Behavioural Sciences

30 August 2013
Thesis advisor: Prof. dr. R.D. Gill

Mathematisch Instituut, Universiteit Leiden

Contents

1 Introduction

Statistical approaches to analysing evidence play a central role in the criminal justice system in the 21st century. Credible approaches to statistical analysis and their implementation in the court of law are therefore powerful entities in everyday life. When applied appropriately, the mathematical modeling of statistics can uniquely enhance the justice system. When applied inappropriately, it can undermine the justice system entirely. In this thesis, a method is developed that can be applied on data that is coming from mobile phone records, which can potentially assist law enforcement agencies in their criminal investigation. During the past two decades, the development of mobile phones has progressed incredibly fast. In 2012 it was reported that 75% of the world's population has access to a mobile phone. The number of phone subscriptions has increased from 1 billion in 2000 to 6 billion in 2012 [15]. It has also been noted that often multiple phones are owned by one person, leading to the expectation that the number of mobile phones will surpass the human population. One reason for owning multiple phones could be to cover up criminal activity. Some criminals use a second (anonymous) phone solely to communicate about such criminal activity, in order to make themselves harder to trace. This development forces police departments to find new methods to detect perpetrators by investigating phone records. When it is known that a crime was committed at a certain time and place, phone records will show which mobile phones were active in the area at that particular time. This could assist the authorities in finding the perpetrator or could be used as supporting evidence. However, especially among criminals, phones are often not registered to a known person, but are bought anonymously. By studying phone records, patterns can occasionally be detected that connect two phones to each other. When the activity of two or more phones coincides both in time and location, a plausible explanation is that the phones are owned by one person. When a connection can be made between an anonymous phone and a registered phone, this can be of additional value in a criminal investigation. Within this thesis, a method is developed to test a suspected connection between two phones, and the statistical value of such evidence is examined in an imaginary case for which data is simulated. Furthermore, approaches for distilling the vast amount of available data into credible evidence are implemented and evaluated. This thesis aims to build on the advances made in this exciting field and safeguard against the ever-present and sometimes catastrophic pitfalls that are encountered to this day within the criminal justice system.

1.1 Statistics in court

Over the past decades, the use of mathematical formulas and/or probabilities in court has been subject to debate, and is still much discussed within the field of forensic statistics. In 1970, Finkelstein & Fairley reviewed the decision of the Californian Supreme Court in the *Collins case* to reject the prosecutor's attempt to link the defendant to a crime by applying the product rule to the probabilities of the relevant characteristics [8]. Multiplying probabilities to calculate a joined probability is only valid for independent events, and the court correctly declined the prosecutor's argument. The use of mathematics in the Collins case contributed to the discussion on whether statistics should be allowed to be introduced in the trial context in order to give weight to identification evidence. Finkelstein & Fairley state that for certain cases, it can be useful to statistically render an expert's opinion, in order to assess the significance of the data [8]. They further argue that *"it is appropriate to translate frequencies of such events into a probability statement by combining them with prior probabilities through the use of Bayes' Theorem"*. One must always be aware that jurors can be swayed by a mathematical explanation of events. Subsequently, inappropriate statistical reasoning could lead a jury to hand down a conviction that is disproportionate to the strength of the evidence that supports it. However, Finkelstein and Fairley feel that mathematics correctly used can be valuable in the evaluation of identification evidence [8]. Tribe (1971) did not share this point of view, which led to an interesting discussion between Finkelstein & Fairley and Tribe [13], [9], [14]. Tribe studies the possible dangers that arise from the use of mathematics in the legal process and finds its alleged usefulness greatly exaggerated. The benefits of allowing mathematics at trial cannot equipoise the severe consequences of possible errors. Tribe concluded that the use of a Bayesian approach, or any similar technique, would be more misleading than helpful and stated that the conjunction of mathematics and the trial process would be more dangerous than fruitful [13]. In reaction to Tribe's attack on their proposal, Finkelstein & Fairley once more emphasize that an expert statistician can assist jurors with the interpretation of statistical evidence that is coming from other experts. Numbers and probabilities are already present in court, since a fingerprint expert will for instance state that the print found at the crime scene matches the defendant's print, and that such prints occur with a frequency of one in a thousand in the population. The expert statistician can help interpret this number and assist on the issue whether the trace is left by the defendant [9]. Finally, Tribe maintains its opinion that the techniques proposed by Finkelstein and Fairley can easily be misused. He cannot think of a reason why a statement of an expert statisti-

cian will help a juror any better to place the evidence in perspective than the one-in-a-thousand finding. Real-life situations are rarely so straightforward that a mathematical model can grasp all its subtleties [14].

Recently, the discussion on statistics in court has resurfaced due to a verdict in a criminal case in the United Kingdom. In 2010, the Court of Appeal ruled that it is accepted to use a mathematical approach to calculate the match probability in DNA cases, based on the reliability of the statistical database [1]. For other areas, such as the identification of footwear marks, data on the distribution and use on footwear is not sufficient to allow an expert to express an opinion based on the use of a mathematical formula. *"Outside the field of DNA (and possibly other areas where there is a firm statistical base), this court has made it clear that Bayes theorem and likelihood ratios should not be used"* [1]. Fenton & Neil (2012) reacted to this ruling by stating that a proper use of probalistic reasoning can potentially improve the efficiency and quality of the criminal justice system. *"It is irrational to assume that some forensic evidence is statistically sound, whereas other less established forensic evidence is not"* [6]. Critics of the ruling further recognize flaws in the current presentation of probalistic evidence, but nevertheless convey their concerns on the impact of the ruling for future evidence presented by expert witnesses. Berger, Buckleton, Champod, Evett & Jackson (2011) state the following: *"the evaluation of evidence for a court of law is not just a matter of "using likelihood ratios" but one of working to a set of principles that are founded on logic. To deny scientists the contemplation of the likelihood ratio - whether quantitative or qualitative - is to deny the central element of this logical structure."* [3]. When the statistical value of a new kind of evidence is evaluated, as presented in this thesis, it is good practice to be mindful of current discussions pertaining to the use of statistics in court.

1.2 Judicial miscarriages

When mathematics and probabilities are used for the purpose of a criminal case, it is of utmost importance to take heed of fallacies that have led to miscarriages of justice in the past. In the evaluation of legal evidence, ideally the circumstances would be such that not only the expert witness, but also the (prosecuting and defence) lawyers and judges possess sufficient knowledge to understand the numbers presented in court. Unfortunately, over the past few decades, miscarriages of justice have still occurred due to the use of statistics in court. Schneps & Colmez (2013) [11] discuss several criminal cases where wrongly applied statistical methods or incorrect interpretations of the calculated probabilities played an important role in the final verdict. Mistakes

made in court can lead to the conviction of innocent people, with severe consequences for their lives. Two recent examples of judicial miscarriage due to the use of statistics will be briefly discussed.

1.2.1 Sally Clark

In the United Kingdom, Sally Clark experienced the effect of wrongly calculated probabilities. The loss of her two baby boys by an unknown cause (also known as crib death) was considered as two independent events. The (low) probability of experiencing a crib death in her social class was squared, without taking into account that it is very likely that these two boys did not die independently and that there was an underlying undetected genetic disease or disorder. The outcome was disastrous as Sally Clark was depicted as a ruthless child murderer, spent years in prison, never recovered from this wrongful accusation and died from an alcohol overdose four years after her release. For Sally, the event of losing a second child to an unknown cause should have never been considered to be independent from the event of losing a first child. Nevertheless, this calculation was made, and the result has been tremendous [11].

1.2.2 Lucia de B.

In the Netherlands, the case of Lucia de B. is well known when it comes to wrongfully applying probabilities in court. This dramatic course of events started when colleagues of this nurse noted that she was often present during an unexpected death of a baby, causing her supervisor to come suspicious that her presence had something to do with the deaths. Data was collected on the number of shifts with or without incident, with or without Lucia present, and a statistic was assigned to the probability that the incidents among her patients had happened under her supervision by chance. From this data, a Dutch psychologist (who was assigned as an expert witness) concluded that this combination of events would only occur by chance in about 1 of 9 million cases of a nurse working at a hospital for nine months. This number was multiplied with the probability calculated from similar statistics coming from the two other hospitals where Lucia worked, leading to a baffling number of 1 in 342 million. Since this number is so small, the conclusion was drawn that it simply could not have occurred naturally. With help from the media, Lucia was soon painted as a remorseless baby killer, and she found herself suddenly accused of multiple murders and attempted murders. However, after her conviction and a reopening of the case, it was demonstrated that the numbers in the dataset that documented Lucia's presence and the

occurrences of incidents were incorrect in the first place, and that the multiplication of the probabilities should have never been done, since the events cannot be considered independent. In 2010, Lucia de B. received the verdict of not guilty, after spending more than six years in prison [11].

The examples mentioned above show that even in the past decades, misunderstandings are still present when numbers and probabilities are used in court. Both cases suffer from an inappropriately applied multiplication of probabilities of events that cannot be considered as independent occurrences. As mentioned previously, the statistical value of a particular type of evidence will be addressed in this thesis. Mathematics is used and the probability that a certain event occurred by chance is central to this case. With the recent miscarriages of justice in mind, the presented experiments will be related to the current topics in the field of forensic statistics. Throughout the thesis it is elucidated which errors should be considered as having the potential to recur in future when a statistical value is assigned to legal evidence which can be presented in court.

1.3 Applicability

The method developed in this thesis can be applied to test the hypothesis whether two phones are probable to be used by the same person. When phone records are available of two phones that show a similar pattern with respect to the time and location of their calls, this method can be used to test the significance of this cohesion. Within the field of law enforcement, this can be of use in criminal cases where a phone is found to behave in a suspicious way around the time of the crime. If the phone is related to the crime, it probably has no contract attached to it from which the identity of the suspect can be detected. It is possible that the perpetrator is using this anonymous phone to prepare or execute the crime, and actively uses a second (registered) phone. It is easy to imagine that a criminal who is using a phone in the preparation of a criminal conduct, is carrying another phone to use for day-to-day and personal business. The phone records of these two phones will show similarity in the time and location of their calls. The method presented in this thesis aims to determine the significance of this observed connection between phones, in order to potentially increase the reliability of presented evidence in court. A defence argument could be that a search among all phone records will yield a connection between phones by chance and that this does not necessarily mean that phones are used by the same person. The current study will therefore discuss a method to quantify the evidential value of such evidence.

1.4 Cell Site Analysis

When the physical movements of a mobile phone are reconstructed, one enters the field of Cell Site Analysis. Mobile telephone operators store call data that can be analysed to determine the location of a mobile phone at the time of a call, SMS or MMS. Analyses on these call data records can produce evidence that can possibly be of use in a criminal investigation where a suspect's movement is of interest. An alibi could be checked, movements of suspects could be followed or the presence at a crime scene could be confirmed.

The method introduced in this thesis can contribute to the field of Cell Site Analysis when one wants to attribute a phone to an individual through the use of Cell Site Analysis. When a phone has no contract attached to it, it is difficult to attribute an individual to it. When there is a phone without contract of note in the investigation, the movements of this phone could be compared with attributed contract phones, to see if a connection can possibly be detected between the anonymous phone and a contract phone. Dr. Iain Brodie states that such analysis on "clean and dirty phones" is a common method used in law enforcement [4]. Focussing on time and location, dr. Brodie looks for similar patterns of movements for the two phones and tries to find a sufficient amount of similarities in order to show beyond reasonable doubt that two phones have been moving together. It is thereby necessary to analyse thoroughly at different times and on different days in order to produce convincing evidence. The method developed in this thesis can be of additional value in cases where the movements of two phones are compared to see if they have possibly been used by the same person, since it aims to statistically test the evidential value of such evidence.

1.5 Terminology

Colocation analysis is seen as the science of determining, from mobile phone records, whether or not two different mobile phones are actually being used by one and the same person.

Dislocation is seen as the occurrence of two calls on two phones close in time but distant in location. This event will also be called a *dislocation event* or an *anomaly*. We state that *dislocation has occurred*, or that two phones have *dislocated*. Call records obtained from providers of mobile telephone servers often only show the location of the cell tower and the specific sector (a range of directions), and not the actual location of the phone making a call. When it is only possible to work with a (fairly good) indication of the

particular area in which the mobile phone was located at the time of the call, it is necessary to approach the claim of a dislocation event with caution. Therefore, the definition of anomaly (dislocation) should be very stringent. Two phones should only be defined to dislocate when it is absolutely obvious the two phones are not in use by the same person, considering the (large) distance between two towers and the (small) difference in time between two calls.

Colocation means that any dislocation events are absent for a long period of time. Note that colocation is defined in a negative way. It can be stated that two phones *colocate* over some period of time if during this period they are never used so close together in time at two locations so distant from one another, that one can rule them out being carried by the same person.

Activity will be introduced as a measure to indicate the intensity or frequency of use of the phone. This measure increases when the number of calls made (and/or received) by the user of the phone goes up.

Mobility will be introduced as a measure to indicate how mobile the user of the phone is. The higher the cumulative distance between the calls of a phone, the higher the mobility of the user.

1.6 Aim of the research

The aim of this research is to determine to what extent colocation analysis provides evidence of the identity of the users of two phones and to quantify the strength of this evidence. To achieve this goal, two different investigative methods are proposed.

1.6.1 Discovering colocating phones

When Cell Site Analysis is used to discover colocating phones, the patterns of movements of phones are compared, and it is analysed if phones show any dislocation event over the period of time in which both phones are active. The first part of this thesis proposes a method to study the probability of dislocation for phones carried by two different persons. This will become an explanatory analysis, to gain knowledge on the overall time till dislocation between phones that are actually not moving together. A study design is proposed to investigate whether the chance of dislocation remains constant over time, and how it is influenced by activity and mobility. With this, it can

be studied whether each new day gives an independent chance of dislocation and whether this chance stays constant, for constant activity and mobility.

Possible research questions *Does dislocation have a half-life and what is it?*
In what way does the mobility and activity affect the half-life of dislocation?

Hypothesis Consider two phones belonging to different persons which have colocated over a short period of time. Then provided their mobility and activity remains constant, their time to dislocation exhibits a constant hazard rate, in other words is exponentially distributed.

It is expected that dislocation has a half-life and that the evidence provided by colocation is stronger (a) if it continues for a longer time period, (b) if the activity of the phones increases, and (c) if the users of the phones are more mobile.

1.6.2 Evaluating colocating phones

The second part of this thesis explores the possibility of statistically quantifying the closeness of the paths followed in time and space by two mobile phones. When phones only colocate over a short time period, but show very similar behaviour in their use over time and the corresponding locations it is difficult to say anything conclusive about their connection. Statistical methods are explored to quantify the strength of colocating evidence, so that the conformity of the phones can be statistically tested.

Possible research question *In what way can permutation methodology quantify the closeness of paths followed in time and space by two mobile phones?*

It is expected that with use of permutation methodology, the visual impression about the closeness of two paths of phone records can be given a statistical value.

2 Discovering colocating phones

This analysis emanates from the hypothesis that two phones carried by two different persons have a certain chance of *not* dislocating in 1 day. Suppose this chance is equal to 50%. One might expect that one day later, the chance of still no dislocation having been observed will have been halved to 25%. After ten days the chance is less than 1 in a thousand, after 20 days less than 1 in a million (2^{20} = 1,048,576). The probability that dislocation *has* occurred is the complement of the chances just mentioned, and exhibits an exponentially fast approach to certainty. If the probability to dislocate remains constant over time, it is possible to designate a fixed length of time such that after each time period the chance of still no dislocation halves, also known as *half-life*. Such consistent behaviour can only be expected after controlling for all factors which influence the chance of dislocation in one day. For instance, if two phones are used infrequently, they will have a small chance of dislocation per day. If two phones *do not* dislocate for a long time, yet are carried by *different, unrelated* persons, then those two persons apparently have very similar, regular behaviour. For instance, they always make the same train journey every day and only use their phones on the train. Then, given they didn't dislocate for a long time, the chance they will dislocate the next day is small. With the presented study design below, it can be investigate whether the chance of dislocation remains constant over time, and how it is influenced by activity and mobility. Is it indeed the case that each new day gives an independent chance of dislocation and does this chance stay constant, for constant activity and mobility?

2.1 Proposed study design

In order to study the probability to dislocate for phones carried by two different persons, data has to be collected from all phone records over a period of time of at least six months. These phone records should consist of the time and location (cell tower section) of calls made. From this data, a phone has to be selected that will become the main focus of the study. Preferably, the phone record of this phone shows mediate activity and mobility. From now on, this (imaginary) main phone will be referred to as phone X.

Before it is further elaborated which information should be additionally collected from the available data, three key notions will be defined. A **pattern** occurs when phone X is active within 4 days at 3 locations that are more than 5 kilometres apart. The three locations and corresponding points in time of the calls that were made, are called the three **points** of the pattern.

Table 1: Proposed design of the data - imaginary values

Pattern	Hits	Phone	Anomaly	Times	Locations	kilometres	After pattern
1	3	123456	yes	27	9	101.78345	-
1		Phone X	-	22	12	67.18456	-
2	2	987654	no	8	3	10.87458	1
2		Phone X	-	3	3	4.23749	-

Furthermore, a **hit** occurs when a second phone is found to colocate with phone X at all three points of the pattern. The colocating phone can be referred to as the hit. Colocation at a point of the pattern is defined to exist when the time of the colocating call is within an interval of 15 minutes before and after the call by phone X, and when the location of the colocating call is the same cell sector, or a cell sector within a 1 kilometre radius.

For each pattern of phone X, the following information has to be collected: what number of phones can be considered a hit and what are the corresponding phone numbers. For phone X, as well as for each hit, information has to be gathered "within" and "after" the pattern. It is suggested that per day within and up to four days after the pattern data is collected on: the total number of calls involving that phone (incoming and outgoing calls), the total number of cell towers activated by the phone and the accumulated distance (cumulative distance between activated cell towers, in order of activation). Also for each day it has to be noted whether or not the phone records indicate an anomaly between phone X and the hit. Finally, if no anomaly occurred within four days after the pattern, it has to be examined on which day thereafter the first anomaly was observed. Table 1 displays a fictitious example of the structure of the required data [1].

2.2 Methods

When a study is performed in the way described above, hopefully many hits are generated. From the collected information on these hits, the time till dislocation can be studied. One can imagine making a plot in logarithmic

[1]The columns "Anomaly", "Times", "Locations" and "kilometres" should be documented per day within and up to four days after the pattern. The last column "After pattern" indicates on which day after the pattern an anomaly is detected, in case no anomaly occurred in the four days after the pattern

scale of the cumulative daily risk of dislocation after the pattern against time. The technical name for this graphic is *the Nelson-Aalen plot of cumulative risk* [2]. The height of the steps will represent per day the estimate of the chance of an anomaly on that day, given that there was no anomaly on the preceding days.

2.3 Stratification

As mentioned previously, this proposed study aims to investigate the relationship between dislocation time and activity and mobility. The users of the phones can be categorized with respect to their activity and mobility, to see the effect on dislocation time. By stratification according to a simple measure of activity and mobility of both the hit and target phone, the role of heterogeneity can be investigated. This categorization could be done by classifying phone X and each hit, at every pattern, as either "high" or "low" activity, depending on whether the total number of calls made during the four days after the pattern is above or below the median for this variable. Similarly, all hits and the target phone at each pattern are scoring either "high" or "low" on mobility, depending on whether the total accumulated distances over the four days after the pattern is above or below the median for this variable. Then, per hit it can be counted how often one of those four binary variables (activity hit, activity target, mobility hit, mobility target) is scoring "high", resulting in an *activity-mobility* score of 0, 1, 2, 3 or 4.

2.4 Example data

In an example data set that is analysed according to the above described study design, it was found that the time till dislocation was heavily depending on the activity-mobility score. Half-lives of approximately 4, 2 and 1 days were found for three (merged) score categories, respectively 0/1, 2 and 3/4. The chance of colocation being maintained for ten half-lives, is the same as the chance of observing heads ten times in a row when tossing a fair coin: $(0.5)^{10} = 0.00098$, just less than 1 in a 1 000. Double this period again, the chance of colocation is less than 1 in a 1 000 000. So even for the lowest score category with a half-life of 4 days, the chance of colocation being maintained for 80 days is less than 1 in a 1 000 000. If this daily dislocation rate of p is maintained for d days, then the chance that no dislocation at all occurs within n days is equal to $(1-p)^d$. Figure 1 shows the graphs of this function for three different values of p corresponding to the three activity-mobility categories defined earlier. For all score categories the chance that colocation

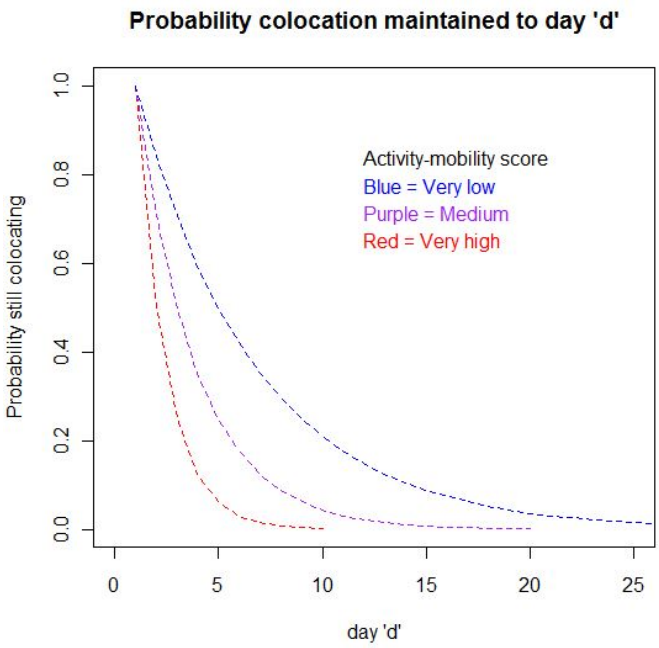


Figure 1: *Chance of no dislocation by day d when daily dislocation rate is constant, by three categories of activity and mobility of phone X and hit. Cut off = 1/1000.*

is maintained decreases in exponential fashion to zero, but how fast the approach to zero occurs depends strongly on activity-mobility score.

2.5 Conclusion

With the study design that is proposed in this section, colocation-dislocation can be empirically studied. To test whether or not dislocation is influenced by mobility and activity, a score can be assigned to each hit along to the amount of mobility and activity of the target and the hit phone. In an example data set it was found that the time till dislocation was heavily depending on the activity-mobility score, and that three different activity-mobility groups resulted in half-lives of approximately 4, 2 and 1 days. From the fact that time till dislocation is found to be affected by the activity-mobility score, one can conclude that it is of utmost importance to take these two features into account in future cases where the evidential value of colocating behaviour of two phones is studied. These findings furthermore imply, especially for hits that are in higher categories, that colocation for a very long period of days occurs rarely by chance.

2.6 Discussion

Within this section, it is discussed which fallacies that have led to miscarriages of justice in the past are relevant, and which considerations have been taken into account in order to prevent the same errors occurring again.

2.6.1 Tunnel vision

When Cell Site Analysis is used to assist investigators in a criminal investigation, one has to be aware of the problem of tunnel vision. Although this fallacy has no relation with the use of mathematics or probabilities in court and can occur in every criminal case, it is important to remain aware of this occurrence when the experiment is actually applied within the context of a criminal case. For the discussed cases of judicial miscarriages in the introduction, tunnel vision was definitely a key factor in the conviction of Lucia de B. and Sally Clark. Both women were depicted as ruthless child murderers by local and national newspapers, which probably made it hard for the judges to look at the evidence in an objective way. Findley & Scott (2006) state that tunnel vision can lead to depraved consequences when it evolves in the criminal justice system. They define this occurrence as a *"compendium of common heuristics and logical fallacies, to which we are all susceptible, that lead actors in the criminal justice system to focus on a suspect, select and filter the evidence that will build a case for conviction, while ignoring or suppressing evidence that points away from guilt"* [7]. When investigators use telephone data in their search for a suspect, tunnel vision can certainly be an issue. Once the investigators have their sights focussed on a person, it is always hazardous that every favourable finding for the defendant is interpreted differently and revolved into something disadvantageous. It is therefore important that an expert witness always questions the investigation methods and takes a critical look at the path that is followed to arrive at a suspect.

2.6.2 Unlikely events are not impossible

Within statistics it is very common to report the p-value, and a low p-value is often accepted as a ground on which one may reject the hypothesis that the events occurred by chance. A p-value under 1 in 1000, is generally approved as a legitimate reason to question whether the hypothesis of pure chance can be rejected. However, one should bear in mind that occasionally very unlikely events do occur in this world. Think about Sally Clarke, who lost two baby boys to an unknown cause. An occurrence that is highly unlikely to take place did occur. Findings that can be interpreted as highly unlikely can therefore be a motivation for further analysis, but one should be careful in stating

definite conclusions. Further, a low probability for an event to occur, should not be confused with the probability of innocence. For both Sally Clark as Lucia de B., the public interpreted the (incorrectly calculated) extremely low probability for the events to happen naturally, as the probability that the defendant might be innocent.

2.6.3 Large dataset

In order to execute the proposed study design, data must be available on many phone records. One is probably working with a data set that can be considered very extensive. When a search is done in a large dataset, one should keep notion of some peculiarities that can take place with respect to the random match probability. This issue often arises in cases where a trace of DNA is found at a crime scene and it is attempted to calculate the probability that a defendant randomly matches the found DNA. A case where a suspicious phone record is found around the time of a crime, deals in a way also with a found trace at a crime scene, and investigators will have to try to find a match. However, one should keep in mind that even an extremely low random match probability can lead to some random matches when a search is performed in a very large data set. The amount of random matches you expect can be counterintuitive. When you come across a random match probability for a certain event of one in 13 billion and a search is performed among 600000 people, you already expect 13 random matches, since the number of comparisons you make is $N(N-1)/2$, which in this case equates to 180 billion. However, when a profile is already present at the start of the search, the number of expected random matches is equal to the random match probability multiplied with the size of the dataset. With the proposed investigation methods, a suspicious phone profile is detected and after that a search is performed among the phone records for a match. To expect a counterintuitive number of random matches is not the issue, but with the magnitude of the data set, one should still be careful when interpreting random match probabilities of one in a million.

3 Evaluating colocating phones

As mentioned previously, the second part of this thesis explores the possibility of statistically quantifying the closeness of the paths followed in time and space by two mobile phones. The method developed in this thesis can be applied to test the hypothesis whether two phones are probable to be used by the same person. It is probable that phones that are used by a criminal will only colocate over a short period of time, making it therefore difficult to say anything conclusive about their connection. Within this section statistical methods are explored to quantify the visual impression of colocating behaviour. To introduce the methodology, data is simulated on the behaviour of a person moving around and making phone calls.

3.1 Texas sharpshooter

Before the methodology is further explained, it is important to weigh in mind whether the selection of two phones for comparison is not facing the so called Texas sharpshooter fallacy. The name is derived from the story of a Texan farmer, who painted a target on the side of his barn around a previous shot bullet hole. Afterwards, he was considered a competent sharpshooter, an obvious fallacious reasoning. Within a research or investigation, the problem of Texas sharpshooter often arises when a pattern is presumed to exist between pieces of data that actually have no connection with one another. Thompson (2009) studied the occurrence of the Texas sharpshooter fallacy within the forensic field and concludes that *"DNA analysts tend to underestimate the frequency of matching profiles (and overestimate likelihood ratios) by shifting the purported criteria for a 'match' or 'inclusion' after the profile of a suspect becomes known"* [12].

When the closeness of two telephones in time and location is tested, the probability is calculated that a certain event occurred by chance, which is similar to research questions in DNA or other forensic matching methods. When the broader context is disregarded and one is intensively focussed on a particular outcome, calculations like this can be fallacious. One has to be careful that no significance is assigned to random results by considering it retrospectively in an exorbitantly narrow context. When an investigative unit starts their search for colocating phones without specifying hypotheses prior to the gathering of the data, and questions the significance of the found patterns after they are detected, this can be a hazardous consecution of proceedings and one could question such a procedure. It would be illogical to test a hypothesis on the same data that was used to generate that hypothesis. However,

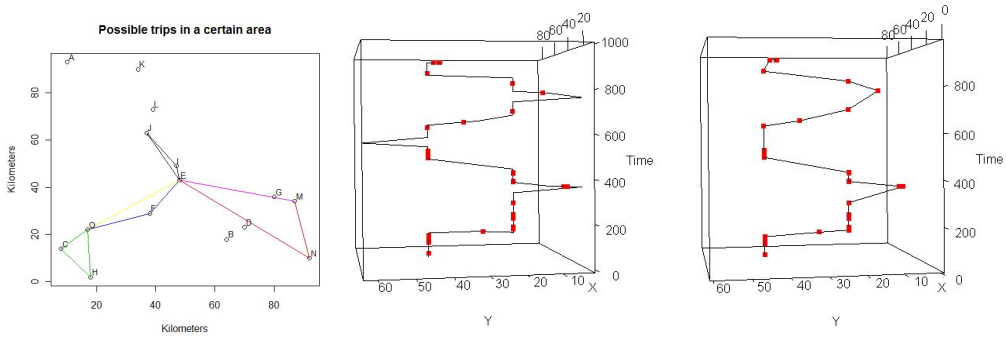
when the study is executed that is proposed in the previous section, phones are included in the study based on patterns, covering at most four days. Following up initial hits in time allows the researcher independently to decide whether or not the initial pattern of colocation is due to chance. When hits are followed for a few days and many phones seem to dislocate, that means that a small part of the data already leads in the direction of longer colocating phones. The findings of half-lives of 1, 2 and 4 days suggest that an explanatory analysis based only on a small bit of the data can already lead investigators towards colocating phones. The methodology introduced in the current section can be applied on a separate, independent part of the data, with many more subsequent days. The statistical analysis on an apparently colocating pair of phones can therefore be seen as an independent confirmatory analysis on a long time history. Thus, there is no need to worry about the Texas sharpshooter fallacy.

3.2 Aim

When records of the calls of two phones are obtained, it is possible to visually represent the path of the two phones both in time and location. In a large data set consisting the phone records of many phones, it could be found that the pattern of times and locations of the calls of two phones together is consistent with the two phones being carried by one person. This similarity could of course be a coincidence, but this study aims to introduce a method that can test whether or not two paths followed by the two phones are so close together that the hypothesis that they are carried by two different persons can be ruled out. In other words: how large is the statistical likelihood of chance colocation?

3.3 Visualisation

Before the method is further explained, it is important to obtain a solid understanding of the idea behind the developed methodology. With a number of graphs it is attempted to visualise a person moving around making phone calls and to envision what it looks like when two phones are colocating for a period of time. In order to make these graphs, a little data set is simulated where we imagine a person moving around throughout an area of 100 by 100 kilometres, commonly visiting 15 locations within that area. Imagine this person moving not randomly, but along a fixed collection of trips. For an imaginary individual, a pattern of 30 randomly chosen consecutive trips is simulated from the trips that are displayed in Figure 2a. Trips where he stays at the same location for a while are also included.

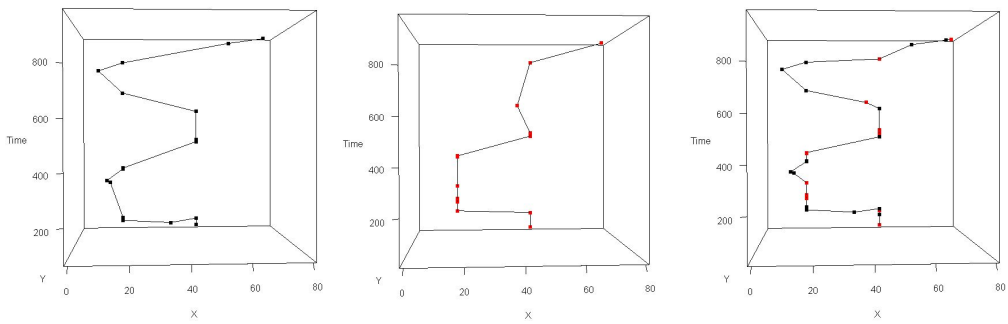


(a) Imaginary map of an area in which a person moves around according to trips

(b) Pattern of movement for an individual

(c) Pattern of movement for a phone

Figure 2: *(b) and (c): Difference between person moving around and making phone calls and the pattern of movement that is visible from his phone*



(a) Pattern of movement for a phone

(b) Pattern of movement for a second phone

(c) Pattern of movement if the two phones were used by the same person

Figure 3: *Visualisation of the difference between unique patterns and patterns put together*

Assuming a traveling speed of 60 kilometres per hour, it is possible to produce a 3d plot of the path he has followed, putting the location of the individual together with the time he would be there. During his movement he might make some phone calls, indicated with a red dot in Figure 2b. However, exact information on the whereabouts of individuals is not available when one is working with phone record data. Figure 2c displays the pattern of the same individual as Figure 2b, but then by only taking the location of his phone calls into account. Thus, when the location (longitude and latitude) of the cell tower that picked up the signal of the calls is taken into account together with the time of the calls and subsequent calls are connected by straight lines, it is possible to produce 3d plots of the path a phone has made, or even of the path a pair of phones has made.

As mentioned before, the methodology developed in this thesis is built upon the assumption that people will occasionally carry two phones around. Imagine a business man who wants to separate business from his personal activities, or a criminal who uses an anonymous phone in the preparation of a crime. In such cases, one could encounter phone patterns as displayed in Figure 3a and 3b. At first sight, these patterns do not necessarily look very similar, but when the two phones are put together on a string one observes a phone pattern that is consistent with what a phone pattern would look like if the two phones were carried by one person. Figure 3c shows the path that would have been followed if the two phones were used by the same person. The line connecting the points follows the time sequence of the combined record of calls. The visual impression of this 3d plot is that the red and black points fit together rather nicely onto a single path. When both phones are active, they are in the same location. Occasionally one of the phones "travels" to another area, without the other phone showing activity. The two phones are never making a phone call far apart in space but close together in time. If the two phones are in the same hands, every move that would have been made is perfectly feasible. The method that will be further explained in this chapter, aims to determine the significance of this observed connection between phones, in order to potentially increase the reliability of presented evidence in court.

3.4 Data simulation

To introduce the developed method data is simulated for a person moving around and making phone calls with two different phones over a period of two months (61 days), in an area of approximately 50 by 100 kilometres.

Let's call him John. With the model that is used in the data simulation, it is attempted to make John's profile as realistic as possible. Several aspects were taken into account:

- It is likely that a person who is moving around will frequently go back to the location he lives. Therefore, a location is chosen that he will visit frequently and where he spends most of his nights. John lives in location E and when it gets later than 11.00 pm, he is more likely to travel back to location E.

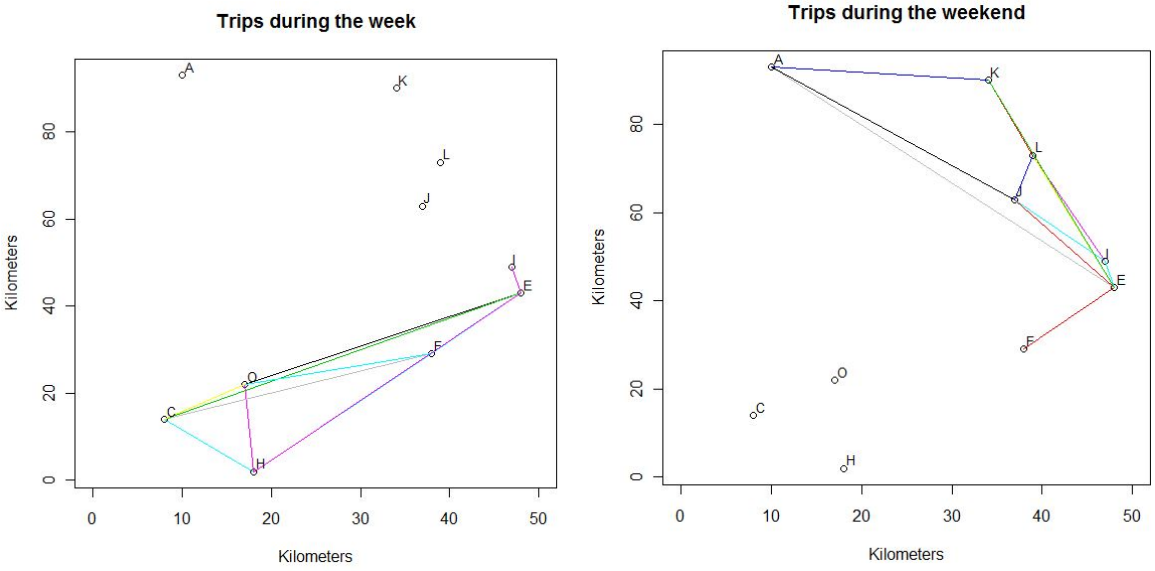
- Figure 4a displays John's possible routes during a week day. The idea is that John has a weekly structure in his movements due to his work, on a work day he moves around locations that are southwest from his home (locations C, O, H, F). From this four week locations, John visits location O most often. With the selection of his next move, there is also a small probability that he travels to locations that he mostly visits during the weekend, but 9 times more weight is given to trips that are in the direction of his work.

- During the weekend John mostly travels in a northwest direction. It makes sense that a person travels to different locations during the weekend than during the week. Perhaps he visits family and friends, or there are some entertaining activities that he likes to do on a day off work. Figure 4b displays John's possible movements during the weekend. Throughout the weekend there is also a small probability that he will visit the locations that are part of the weektrips, but again, with the selection of his next move, much more weight is given to the weekendtrips. On top of that, from all possible locations that are part of the weekendtrips, more weight is given to location K, his favourite weekend location.

- During the night-time, between the hours of 0.00 and 8.00 am, John stops moving and stays in the same location. Most of the nights he stays in location E, but occasionally he spends the night somewhere else. During the night the calling frequency is very low; there are just a few phone calls made in the night-time over the period of 61 days.

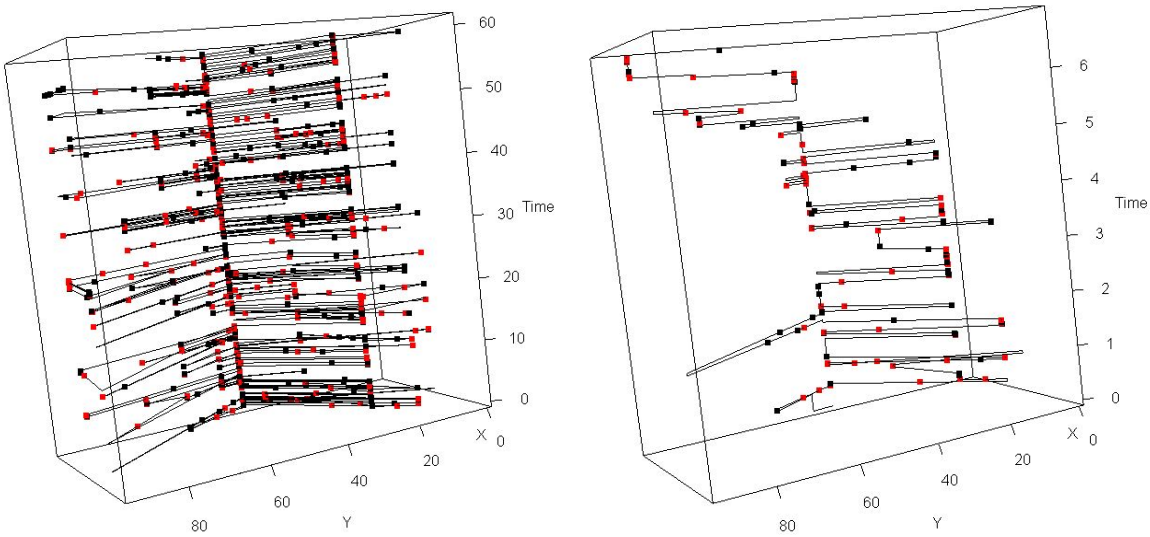
- As mentioned before, John is carrying around two mobile phones and uses them equally frequent. During the day, he switches randomly from one phone to another.

The simulated movements of John and the phone calls he makes are displayed in Figure 6a and 6b, where the second graph shows the same sequence



(a) (b)

Figure 4: *Visualisation of possible weektrips and weekendtrips*



(a) Two months (b) Zoomed in on the first week

Figure 5: *Visualisation of the simulated movements and phone calls over a two month period*

of movements and calls, but then zoomed in on the first week of the 61 days period. The colours of the dots indicate which phone he was using. The dataset with the time and location of the calls of the two phones will be used to explain the methodology that is developed to test the closeness of two patterns of phones. The data is now simulated in a way that the phone calls of the two phones are actually simulated from a sequence of movements of one person. It is important to keep notion of the fact that criminal investigators will separately come across the phone records of the red and the black phone. When they put the two phone records together, they show a path that could be travelled by one person, from which the suspicion arises that the two phones might belong to the same person.

3.5 Methodology

When investigators come across two phones of which they suspect that they are used by the same person, it would be ideal if they can create a large sample of pairs of different persons, who are independently making similar movements and using their phone in a similar way. With such a sample it would be possible to compare the paths of the two phones to a large number of phones that are carried by different persons. Defining these notions in order to collect such a sample from the available phone records can be a difficult task that will cost a lot of time and money. However, when data of only two phones is available, "time" provides us with a unit of repetition with which we can construct an artificial sample of phone histories of similar phone users.

3.5.1 Permutation tests

In general, a permutation test can be used to obtain the distribution of a test statistic, by calculating all possible values of this test statistic under rearrangement of the labels of the observed data points. This test is often applied to test the hypothesis whether or not two different groups of persons or objects actually come from the same distribution. For two groups a certain test statistic can be of interest (for instance the difference of the two sample means), and when this test statistic in the original grouping significantly differs from the test statistics of all possible rearranged groups, the conclusion can be drawn that the original group label matters and the null hypothesis that these groups come from the same distribution can be rejected.

When the paths of two phones are compared, there is no question of comparing two groups of people, but the permutation methodology can still be

appropriate for the case at hand. The path that is formed by the records of a phone can be seen as a collection of days, and for this experiment the days within a path will be shuffled in order to construct an artificial phone history. The order of phone calls within a day will be retained, but the days will be rearranged. The path that is created of a "new" phone, can be seen as coming from a phone that makes similar movements as the original phone, uses the phone equally intensively, spends the night at the same location, and so on. With regard to the two phones of interest, one questions the fact if they are used by different persons, but when a comparison between the path of one phone with a shuffled path of another phone is made, this can be seen as a comparison between two phones that are known to be in in different hands.

How well the two telephone records fit together, can be seen as the statistic of interest. Defining the fit of two mobile phones, and determining their "closeness", will be discussed later in this section. If two phones are carried by different persons, it is expected that rearranging the days of one phone, will not change the closeness of the two phones. In other words, if two phones are not in use of the same person, and the similarity in the paths only arises by chance (for instance because two people live in the same building and work in the same area), shuffling the days of the path of one phone should yield a path that is still equally similar to the path of the other phone. If rearranging the days of one phone affects the test statistic (the closeness of the two paths) significantly, it can be concluded that the original closeness of the two paths is special, and a logical explanation for this could be that the two phones are in use of the same person. Thus, when the actually observed closeness is much smaller than the closeness of almost all the shuffled paths, we conclude that the observed closeness is highly improbable under the hypothesis that the phones are in different hands.

3.5.2 Measuring closeness

The statistic that is used to measure the closeness of two paths is evidently a point of interest that can be discussed. Before this is further elaborated, it must be noted that different kinds of closeness between paths of phones can be detected. In the first case, one can imagine that a person is carrying two phones for two different activities (for instance business and private) and rigorously separates these activities into different time periods every day. Then, the person will use one phone for a period of time during a certain activity, and then switches to the other phone, putting the first phone down for a while, then switches to the first phone again, and so on. When the two

paths are put together, this kind of use of two phones will result in a pattern that has activity of one phone combined with a lull in activity of the other phone and vice versa. The two paths will link up at the switches between the two activities. On the other hand, one can also imagine that a person is carrying two (or maybe more) phones, and uses them both during the same activity, but for communication with different groups of contacts. Thus, the use of one phone will go together with the use of another phone. When the two paths are put together, this kind of use will result in a pattern of activity of both phones, followed by a lull in activity of both phones, followed by activity of both, and so on. This difference is important when a measure of closeness is developed.

To compare the paths of two phones and measure how well the patterns fit together, the distance that is covered every time a switch is made on the same day between one phone to the other is calculated. A *switch* occurs whenever two subsequent calls on the same day in the combined call records of two phones are made by two different phones. Over all the days, these distances are accumulated and provide a measure of closeness of the two paths. A second statistic that is studied to determine how well the paths of two phones fit each other is the number of anomalies that occur when the paths of two phones are combined. As stated in the introduction, an anomaly is defined as the occurrence of two calls on two phones close in time but distant in location. An anomaly is said to have occurred when two calls are made within 15 minutes of one another at distances (of the respective cell towers) of at least 25 Km.

3.6 Results

1. In this section results regarding the average accumulated daily distance are discussed. Recall that the sum of all distances between the two phones is added up over all switches, within the same day, between the two phones. The two phones will be referred to as the *black phone* and the *red phone*, according to the colour of the dots that represent their phone calls.

    - Actual: for the 61 days, the sum of all distances between locations of phones when a switch is made between one phone to another is 5192 Km. The average daily accumulated distance between the locations of cell towers involved in subsequent calls by each of the two phones is around 85 Km.
    - Shuffled: a permutation test is applied, whereby the match of the

actual history of the black phone is matched to the history of the red phone after a random permutation of the days. This was done for 1000 random permutations of the days of the red phone. Figure 6 shows a histogram of the resulting two month average accumulated daily distance between the black and the red phone for these 1000 random permutations. Typically the two month average accumulated daily distance between consecutive calls of different phones is about 260 Km, and in most cases between 220 Km and 300 Km.

2. When we take a look at the second statistic to measure the closeness of the paths, namely the number of anomalies, similar results are found.

   - Actual: in the three month period that has been studied, the black and red phone never exhibit an anomaly defined earlier in this chapter. This result is quite logical since the two phones are simulated from the movements of one person. However, also when investigators find two phones of which they suspect that they are used by the same person, no anomalies will be detected in the original comparison of the paths of the two phones, since an anomaly is defined in such a way that it is a strong indicator of the phones not being used by the same person.
   - Shuffled: all of the 1000 shuffled histories exhibit a substantial number of anomalies. Figure 7 shows a histogram of the number of anomalies found in a two month comparison between the black phone history and 1000 permutations of the red phone history. Typically around 90 anomalies are detected, and the largest part of the permutations show between 60 and 120 anomalies.

As seen before, it is possible to plot two phone paths in a 3d box, to see how they fit together and to get an impression about the difference between the original path and a shuffled path. Figure 8 shows (a) the actual path of the black and red phone together and (b) the path of the black phone with one of the 1000 permutations of the red days. It can be concluded that the comparison between the black phone with a permutation of the red phone looks messier, but to see the difference more clearly one can take a look at Figure 8c and 8d. This plot zooms in on (c) the first 2 days of the actual calls of the black and red phone and (d) the actual calls of the black phone on the first 2 days of the three month time period, together with the calls of the red phone on 2 other days selected at random in the two month period (one of

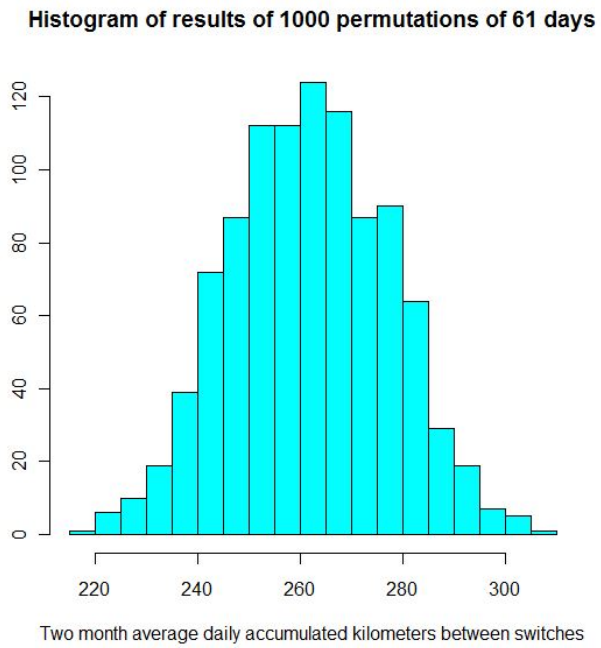


Figure 6: *Daily average total distances between consecutive calls of the black phone and 1000 permutations of the red phone in a 2 month period*

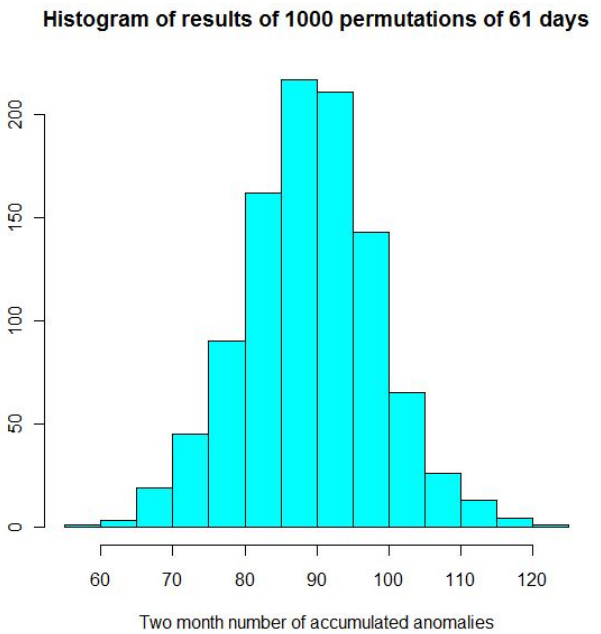
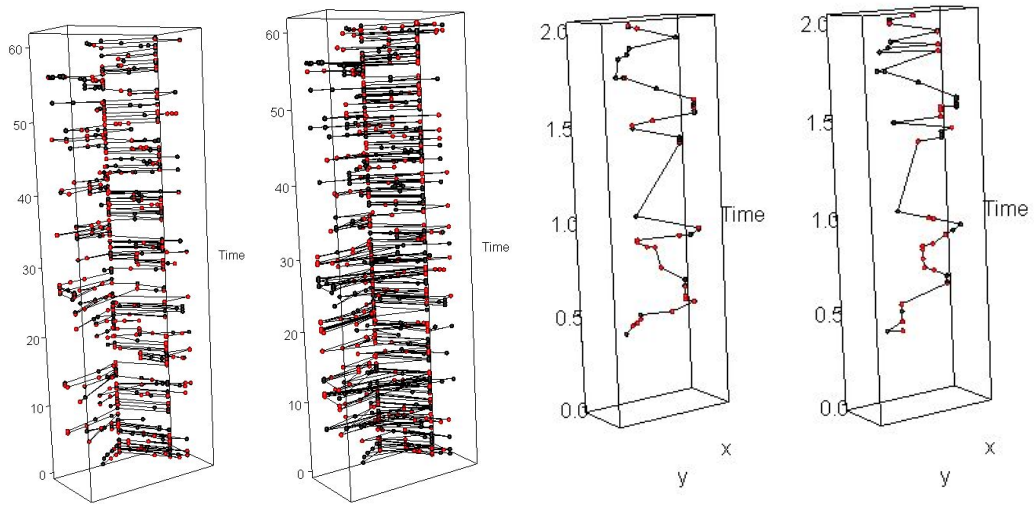


Figure 7: *Total accumulated anomalies between consecutive calls of the black phone and 1000 permutations of the red phone in a 2 month period*



(a) Actual (b) Red days shuffled (c) Actual (d) Red days shuffled

Figure 8: *(a) and (b): Black phone and red phone on a two month period, (c) and (d): Black and red phone on the first two days of the two month period*

the random permutations). Within the actual path the calls of the black and red phone fit together in such a way that it is feasible that they are carried by the same person. Within the shuffled path some strange movements and even some anomalies are detected if the two phones were in the hands of the same owner.

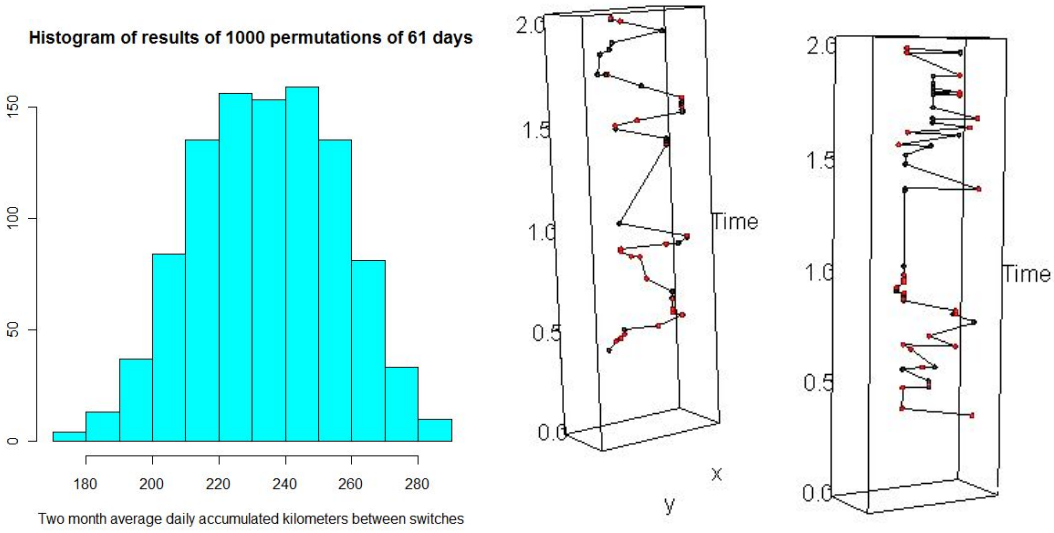
A shuffled, artificial phone record can be seen as a typical history of the calls made by a phone user very similar to the owner of the red phone. As stated before, the *new* phone spends a similar amount of time in exactly the same places as the red phone, it is used just as intensively, it has been taken on the same journeys, and so on. Since it is known that the history of this phone is artificial, comparing it to the black phone results in a combined pattern that one would expect when the two phones are in different hands. Putting the black phone record together with a shuffled red phone record results in clear dislocations (anomalies) on several days, indicating that the original paths of the two phones do show a special connection.

3.7 Reflection on the methodology and an adjustment

Up until now an artificial sample is created of pairs of phones carried by different persons. Within each pair, one phone is similar to the black phone, the other phone is as similar as possible to the red phone. With use of permutation methodology the question is addressed whether the actual black and red combination looks like one of these pairs of phones in different hands, or whether it looks like a pair of phones in the same hands. This is statistically quantified by fixing a numerical measure on the paths of a pair of phones. One can imagine sophisticated measures of the degree of erraticism of the joint path, or equivalently of the degree of mismatch between the two paths. So far a measure of closeness has been studied that is simple and quite directly corresponds to what can be seen when looking at the graphs. Importantly, this measure can be computed easily and rapidly, so it can be computed for many pairs of phones. First, the sum of all distances between the locations of cell towers is calculated between pairs of calls of the two phones which are adjacent in time (and on the same day). If two phones are in different hands, it is expected that from time to time they make phone calls close together in time but in locations which are far apart from one another. On the other hand, if two phones are in the same hands and are used close together in time, there cannot be a large distance between the two locations at which they are used. A second measure that is taken into account is the number of dislocations between the two phones. Both measures are considered in a comparison between pairs of phones (an actual pair, and many artificial pairs formed by shuffling the days of one of the phones) that are similar with respect to the length of time they are followed and their activity and mobility.

3.7.1 Adjusting shuffling

A logical point of critique for this method is that human life shows many patterns and regularities and that the closeness of the paths of two phones is automatically influenced when the days of one phone are randomly shuffled without this patterns taken into account. When measuring the correlation between the locations of calls made by different phones on different days, it is possible that a large correlation between the locations is detected because the owners are making similar movements from one day to another. When the days are shuffled randomly, the extremely strong weekly rhythm to human behaviour is not taken into account. One can imagine that especially in a city, the movements of two persons on week days versus weekend days are different, as is also taken into account when the data of a person moving around was



(a) Only the week days are shuffled with each other

(b) Actual

(c) Red days shuffled

Figure 9: *(a) Daily average total distances between consecutive calls of the black phone and 1000 permutations of the red phone in a 2 month period - week days shuffled; (b) and (c): Black and red phone on the first two days of the two month period*

simulated. When the artificial red path is created by shuffling all the 61 days completely at random, the behaviour of the black phone on some week days is sometimes juxtaposed to the red phone on weekends. Systematic weekly patterns in human movements and behaviour do exist, such that randomly shuffling all the days possibly produces misleading results.

3.7.2 Method

To take care of this problem a new method of rearranging the days of the red phone is implemented. The experiment is repeated, but now all Mondays, all Tuesdays, all Wednesdays, and so on, are shuffled independently, and separately from one another. This way, the distance between calls of the black phone and calls of the artificial red phone are measured on the same days of the week.

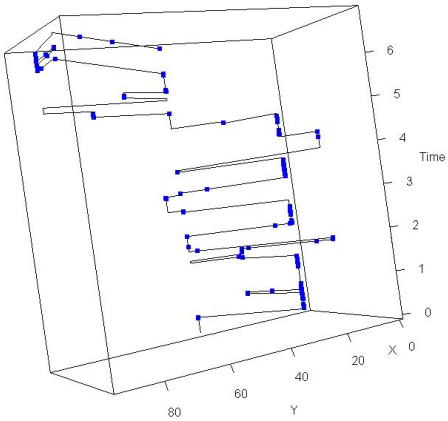
3.7.3 Results

This modified procedure generates smaller average accumulated daily distances between the actual black phone and the shuffled red. However, the two month average accumulated daily distance between subsequent calls of the new artificial pairs is still systematically much larger than the average of the actual pair. Recall that the actual path of the black phone compared with the actual path of the light blue phone, results in an average accumulated daily distance of 85 Km and the average of the 1000 random permutations is around 260. Figure 9 shows the results for a 1000 permutations. Typically the two month average accumulated daily distance between consecutive calls of different phones is about 230 Km, and in most cases between 180 Km and 280 Km. The fact that this procedure generates smaller average accumulated daily distances implies that it might be necessary to take the week structure into account when shuffling the order of the days of one phone.

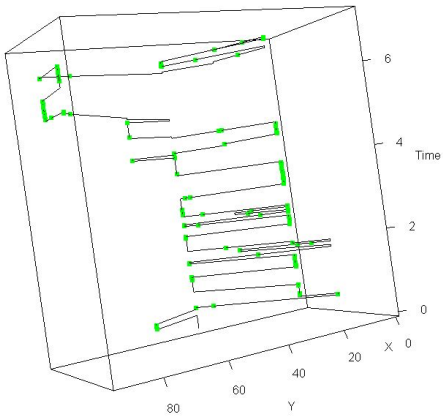
3.8 Validation method

Further experiments have been undertaken in order to achieve two objectives. Firstly, it allows validation of the use of permutation methodology. Secondly, it provides a means to test the assumptions that formed the basis of the adjustment in the shuffling process that has been applied.

Experiment A For a first experiment, data is again simulated on a two month period, but now for two different persons. This way, they can be seen as separate phones that show similar activity, but do not travel the same route. The fit of these two phone records is now tested with the same methodology as used before. The two different phones are from now on referred to as the *blue phone* and the *green phone*, regarding the colour of the dots that represent their phone calls. Within this experiment, the phone records of both persons are simulated with similar assumptions as were stated in John's profile: they both are likely to sleep in E and they travel to similar destinations. A small adjustment was made: during the week, during the day, they prefer to be at location O and during the weekend they prefer location K over the other locations. However, the owner of the blue phone is visiting the weekendlocations on Saturday and Sunday, while the owner of the green phone travels towards that direction mostly on Friday and Saturday. This different week structure is visible in Figure 10, where the first week of the routes are plotted. The blue phone travels mainly towards the weekendlocations on day 6 and 7, the green phone on day 5 and 6. Suppose that despite of this difference, it was suspected that these two phones were

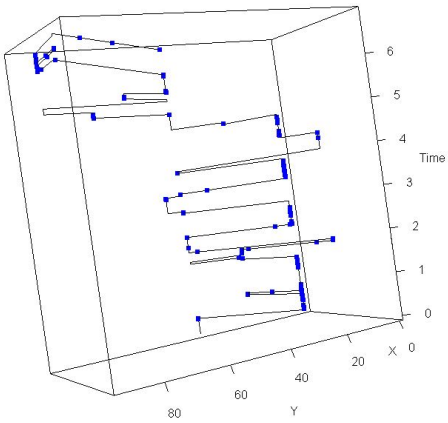


(a) Person 1 - blue phone

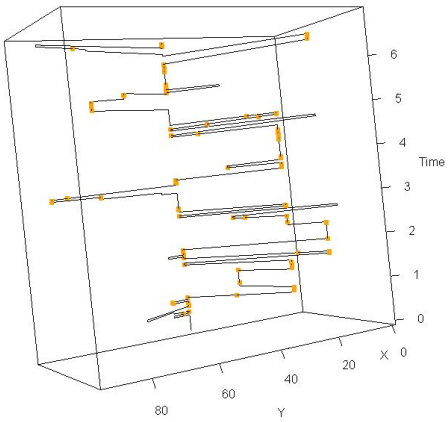


(b) Person 2 - green phone

Figure 10: *(a) and (b): Two persons moving around and making phone calls with a different week structure*

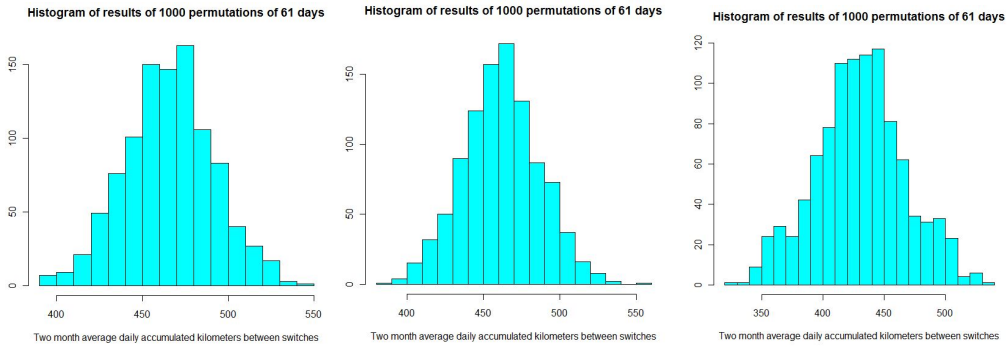


(a) Person 1 - blue phone



(b) Person 3 - orange phone

Figure 11: *(a) and (b): Two persons moving around and making phone calls with the same week structure*



(a) Random shuffle, different week structure. Original: 428

(b) Random shuffle, same week structure. Original: 363

(c) Weekly shuffle, same week structure. Original: 363

Figure 12: *Results experiment A, B1 and B2*

being used by the same person. A permutation test could be applied to test whether the original distance between the two phones is very different from the distance between the blue phone and 1000 permutations on the order of the days of the green phone. Since we know in advance that the phones are actually not being used by the same person, we do not expect the permuted distances to deviate much from the original distance. The same measure of closeness is calculated: the cumulative distance that is covered every time a switch is made on the same day from one phone to the other. Over this two month period, the total accumulated distance is 26108 kilometres. This is a fairly high number, it would mean that a person that is using both phones is traveling more than 428 kilometres per day. The order of the days of the green phone is randomly permuted 1000 times, and for each permutation the measure is calculated again. Around 93% of the permutations result in a higher accumulated distance than the original. A histogram of the results can be found in Figure 12a. This result confirms the expectation that the permuted distances are not extremely different from the original distance that was covered if the phones were being used by the same person. Note that the two owners of the phones in this experiment did not have the same week structure.

Experiment B

1. In the next experiment one of the persons for which data was simulated is now compared to a third simulated person (owner of the *or-*

ange phone), that moves around according to the same week structure. They both travel weekendtrips mainly on Saturday and Sunday, and weektrips during the week. With this comparison, the original distance is calculated over days that have mainly the same week structure, as seen in Figure 11. If one person would use both phones, the original accumulated distance is 22178 kilometres, an average of approximately 363 kilometres per day. Again, 1000 permutations are randomly generated on the days of the orange phone, where these permutations do not take the days of the week into account. A total of 100% of the permutations generate a higher accumulated distance that the original, the histogram of the results is displayed in Figure 12b.

2. With a final experiment, we will try to determine the effect of shuffling the days of the orange phone not randomly, but by the days of the week. Again 1000 permutations are generated on the order of the days of the orange phone, shuffling Mondays only with Mondays, Tuesdays with Tuesdays etc. With this method, around 95% of the permutations generate a higher accumulated distance that the original. Again, the histogram of the results is seen in Figure 12c.

In order to draw conclusions from these experiments, it is important to thoroughly understand the difference between the three experiments. Table 2 to 6 show the original comparison and an imaginary permutation for all three experiments. Note that the permutations of experiment B2 are not random: the difference between the numbers of the days that are compared with each other is divisible by 7.

3.9 Conclusion

Within this thesis, methodology is introduced to test for a pair of phones whether the observed colocation could have occurred by chance. More precisely, with use of this statistical test, the probability can be estimated that a pair of phones similar in all respects to the present pair, but in different hands, would show, by chance, such a pattern of colocation as was actually observed. The first important decision was taken on the measure of closeness of the paths in time and location of two phones, which is a matter of judgment. This measure has to be intuitively easy to understand and easy to calculate on available call records.

The colocation of pairs of phones is investigated, comparing the actual daily accumulated distance between switches with what one would expect by chance for similar phones in separate hands. Data is simulated of one person moving

Table 2: Original Comparison Experiment A

Phone 1	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	Week	Week	Week	Week	Week	Weekend	Weekend
	Week	Week	Week	Week	Weekend	Weekend	Week
Phone 2	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7

Table 3: Possible Permutation Experiment A

Phone 1	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	Week	Week	Week	Week	Week	Weekend	Weekend
	Week	Week	Week	Week	Weekend	Week	Week
Phone 2	Sunday	Monday	Tuesday	Thursday	Saturday	Thursday	Wednesday
	Day 21	Day 8	Day 58	Day 19	Day 41	Day 32	Day 3

Table 4: Original Comparison Experiment B1 & B2

Phone 1	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	Week	Week	Week	Week	Week	Weekend	Weekend
	Week	Week	Week	Week	Week	Weekend	Weekend
Phone 2	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7

Table 5: Possible Permutation Experiment B1

Phone 1	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	Week	Week	Week	Week	Week	Weekend	Weekend
	Weekend	Week	Week	Week	Weekend	Week	Weekend
Phone 2	Saturday	Tuesday	Friday	Thursday	Sunday	Tuesday	Saturday
	Day 20	Day 9	Day 40	Day 53	Day 7	Day 16	Day 34

Table 6: Possible Permutation Experiment B2

Phone 1	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	Week	Week	Week	Week	Week	Weekend	Weekend
	Week	Week	Week	Week	Week	Weekend	Weekend
Phone 2	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	Day 15	Day 51	Day 24	Day 39	Day 61	Day 13	Day 42

around, making phone calls with two phones. Permutation methodology is used to create an artificial sample of phone histories that are similar to the history of one of the phones. The idea is that if two phones are unrelated, mixing up the days of the history of one of the phones generates a new history of a phone of very similar behaviour to the original.

When the time of the calls together with the location of the cell tower sector that picked up the call is taken into account, it is decided to take the average accumulated distance between switches, as well as the number of anomalies, as a measure of closeness. With this data it is possible to make 3d plots, to get a visual impression about the way two telephone records fit together. Data was simulated over a time period of two months of the calls of the black phone and the red phone. From the 3d plot of the actual histories, it can be concluded that the black phone and red phone fit well together on a long string, with no strange movements (no anomalies) if the phones were carried by one person. When the path of the black phone is put together with an artificial path of the red phone, clearly a different picture arises. For the average daily accumulated distance, as well as the number of anomalies, all 1000 permutations show significantly higher values. After a modification of the procedure of shuffling, where the days of the light blue phone are systematically shuffled by the days of the week, the average daily accumulated distance decreases somewhat, but still remains highly significant. Still all permutations give a number about 2 to 3 times as high as the average daily accumulated distance in the actual comparison.

Finally, three experiments were performed to validate the method used to compare phone records. The aim of the experiments was to study the consequences of applying the developed permutation methodology on two phones that are known to be *not* being used by the same person and to determine the importance of taking the days of the week into account when the order of the days of one phone is permuted. To create a pair of phones that are known to be in different hands, two persons are simulated moving through space, both using a phone. First, two persons are generated that travel among the same locations, but have a different week structure. From the results of experiment A it can be concluded that comparing these two phone records that are actually in different hands results in a high accumulated average distance per day, and that 1000 permutations generally do not result in significantly higher distances. Experiment B1 shows that a comparison between two phones that are still in different hands, but move around on the same day of the week, results in a considerably lower accumulated average distance per day. When the order of the days of one of the phones is permuted and the

days of the week are not taken into account, all 1000 permutations generate a higher accumulated average distance. A consequence of this result is that a cautionary note has to be placed with the previously performed experiments that have not taken the days of the week into account. However, even with a random order of the days, the results of the comparison between the black and red phone (that were being used by one person) were more extreme. The average accumulated distance between the black phone and the permutations on red was approximately 2 to 3 times higher than the original, while the average accumulated distance between the blue phone and the permutations of the orange phone is on average 1.25 times higher than the original distance. Finally, from the results of experiment B2 it can be concluded that the accumulated average distance per day for permutations with the days of the week taken into account, is not significantly higher as the distance calculated in the original comparison of the two phones. About 95% of the permutations generate a higher accumulated distance than the original. Since the original comparison consisted of two phones that were actually in different hands, a result like this strengthens the idea that something special is going on when two phones show a much lower accumulated average distance than the distances of all the generated permutations. A probable explanation for such a finding is that the two colocating phones are being used by the same person. Furthermore, the results of the three experiments together imply that it is important to take the days of the week into account when the days of one phone are shuffled.

3.10 Discussion

First, the methodology that is introduced in this thesis will be once more reflected upon. Second, the relationship is discussed between the presented method that can be applied to mobile phone records and other identification methods used in forensic science. Furthermore, relevant research is discussed that studies the identification of mobile phones in a large data set of mobile phone records. Fourth, the relationship is studied between the current investigative methods and commonly applied techniques of Bayes Theorem and likelihood within the field of forensic science. Finally, some ideas for further research are proposed with which the validation of the found results could be further evaluated.

3.10.1 Reflection on the methodology

With the current methodology, a lot of thought has been given to validate the methods used, but since this thesis is discussing newly developed methods,

it is important to continue evaluating the methodology and placing cautionary notes with some of the results. When permutation tests are used, one has to be aware of the possibility that the statistic of interest is increasing significantly because of other factors than the similarity of the original comparison. The results of this thesis for instance imply that it is important to take the week structure that exists in human movement patterns into account when the order of the days of one phone is shuffled. For a comparison of two phones that are not in the same hands, a random shuffle generates average daily accumulated distances that are much higher than the original. A shuffle where one takes the day of the week into account generates reasonably similar distances as the distance in the original comparison, which is a strong indicator that not taking the days of the week into account pushes up the distance measure automatically. However, the found percentage of 95% is still fairly high, indicating that it will be necessary to search for other confounding factors that haven't been thought of. Furthermore, it would be interesting to study if the solution that was thought of to deal with this weekly rhythm is optimal. To see what the results are when for instance not days but whole week blocks are shuffled, would be worth knowing.

A critical note has to be placed with the choice of the distance measure. Up until now it is decided to look at the average accumulated daily distance between switches, but one could think of more advanced methods to compare the closeness of two paths in time and space. It will be interesting to investigate alternative distance measures and see if similar results can be found for different distance measures. As stated in the beginning of this chapter, different kinds of usage of two phones can be thought of. The chosen distance measure can lead to problems when a person is not using his two phones randomly during the day, switching randomly from one phone to the other, but uses them both for a period of time followed by a lull in activity, again for both phones. When a person is carrying two (or more) phones and uses two phones for the same activity, one can imagine that the two phone records seem highly related. Every day, both phones show activity during the same hours of the day and the cell towers that pick up the phone calls of the two phones are not far apart. To be able to statistically quantify the closeness of the two paths followed in time and space in situations like this, it is important to choose a distance measure that will go up when the history of one phone is compared to 1000 permutations on the history of the other phone. However, the original comparison will show a high number of switches per day, since the use of one phone is combined with the use of the other phone. The distances between the switches will not be very large, but a large number of switches could result in a high average accumulated daily distance. Even

more so when one is taking into account that it is possible that a person is staying put at one place, but the phone calls of his two phones are picked up by different cell towers around him, making it look like he is on the move. When the order of the days of one phone is shuffled, this block of activity will be compared to a random other day of the other phone. It is very well possible that this random other day shows activity of the other phone in a different time period of the day. That way there are very few switches, generating a low average accumulated daily distance. A possible solution for his problem is to take the average distance among switches instead of the accumulated distance.

3.10.2 Relationship to other identification methods in forensic science

This section studies the relationship between the new investigative methods that can possibly be applied to mobile phone records and other disciplines concerning identification methods in the field of forensic science. The process of detecting perpetrators by investigating phone records bears much resemblance to other kinds of evidence that are often presented in court. In a way, this investigative method aims to analyse whether a phone pattern can be matched to a person in a similar way a fingerprint expert compares the fingerprint found at the crime scene to the suspect's print. In both situations one is trying to determine the probability that the found conformity is due to chance. Fenton & Neil (2012) introduce a way to address all probabilistic match evidence and discuss common fallacies that affect probabilistic reasoning with regard to evidence. They state that every branch of forensic matching that is based on some properties of people, follow the same underlying principles:

- *"Every person has a 'profile' (defined by the area of forensics) that can be measured by some defined procedure.*
- *In certain circumstances a person leaves a 'trace' (or 'print') of this profile.*
- *In certain circumstances we can measure the profile of the trace that was left.*
- *There is a criterion for determining whether a trace profile matches the profile of a person."* [6]

The calls that a person makes can be seen as a person's profile. Collecting the

phone records from the network providers, including the time and location of all the calls made by a phone, is thereby the defined procedure that measures the profile. A small discrepancy exists between this approach and other areas of forensic matching, since every *phone* has a profile, but a person can carry multiple phones and thus have multiple profiles. A phone record is therefore not matched to another person, but to another phone record. When one of the phones is registered to a known person, it is concluded that the anonymous phone belongs to that person as well.

Furthermore, the phone record of the phone that is found to behave in a suspicious way around the time of the assassination can be seen as a possible trace that is left by the perpetrator. This part of the approach does not diverge much from other forensic matching methods, such as fingerprints. People leave their fingerprints with almost every move they make, but when a fingerprint is found around a crime scene, this can be an indication that the print comes from the perpetrator. People that are frequent callers leave a lot of 'phoneprints' behind as well but when the print is found in such a way that the phone shows activity around a crime, suspicion rises that the profile can be attributed to the perpetrator.

The last point mentioned by Fenton & Neil (2012) is essential with respect to developed methodology. The first part of this thesis proposes a study design to study the likelihood of two phones colocating for a longer period of time, and the second part quantifies the closeness of two paths by the use of permutation methodology. These new investigative methods attempt to define a criterion for determining whether a trace profile matches the profile of a person, or in this case the profile of another phone record. Steps have been taken toward validation of these new methods. However, the reliability of new methods such as these remains open for discussion. Future experiments are required to assess this matter and add credibility to the approach.

Fenton & Neil furthermore emphasize that a match never results in a unique identification of a person. This fallacy often arises in areas of forensics such as DNA matching. It is stated that it can solely be concluded that the profile of the trace is matching another profile within the agreed criteria [6]. With respect to the current methodology, it is important that statements about unique identification are avoided.

3.10.3 Relevant research

Mobile phone records can be used in a wide range of scientific fields. Social sciences can use data on mobile phones to analyse social networks, or within the epidemiological field one may gather information about human mobility. From call patterns of a large number of individuals, a model can for instance be built in order to study the spreading of diseases in a population [10]. No research has been found that uses mobile phone records to determine how long it takes for two phones to show dislocation or to develop a method to study the closeness of two telephone paths in time and location. Studies on telephone records commonly assume that one phone belongs to one person, and no previous research can be found that generates or analyses investigative techniques to find two phones that are in hands of the same person.

A search within the available literature of research that uses telephone records in their analysis did yield one relevant study. De Montjoye, Hidalgo, Verleysen & Blondel (2013) studied human mobility traces by looking at the records of 1.5 million mobile phones over a time period of fifteen months. Data was available on the time of the calls that were made by these phones and the location of the cell tower that picked up the signal. Thus, information was available that is similar to the information of the light blue phone of project two, but then for a longer time period and many more phones. The researchers first altered the data in such a way that not the exact time of the phone call was taken into account, but solely whether a phone call was made in a particular cell tower sector in a timeframe of one hour. From this data it was studied how many random time points were needed in order to uniquely identify the phone. It turned out that only four random points are necessary to determine which phone belongs to those points in 95% of the cases. Thus, when a telephone string was randomly selected and from that four random data points, in 95% of the cases it was uniquely identified which telephone record the points belong to. It was furthermore found that the number of data points necessary to identify a telephone goes up when the length of the timeframe was expanded or when cell tower sectors were taken together [5].

The study of de Montjoye, Hidalgo, Verleysen & Blondel shows some resemblance with the methodology presented in this thesis, but focusses on identifying a telephone and does not study the possibility that multiple telephone records could be used by the same person. Two phones that make a phone call within the same timeframe but in different cell tower sectors are considered to be separate, without taking into account if those two phones show a dislocation in the way that is defined earlier. However, the results

from Montjoye, Hildago, Verleysen & Blondel do show that telephones can be separated from each other rather quick, since only four points are necessary to determine which telephone record belongs to those data points in a very large part of the cases. Although dislocation is not studied, the results do suggest that colocation is not very common.

3.10.4 Relationship to Likelihood Ratio

Within the field of forensic science, research is commonly performed by calculating the Likelihood Ratio. The hypothesis of the defence and the hypothesis of the prosecution are formulated, and the probability of the data under those hypotheses is calculated. In cases where people would like to test whether or not two phones belong to the same person, the ratio would be calculated of the probability of the data under the hypothesis that the two phones are in use of the same person and under the hypothesis that the two phones are used by two different persons. To study this ratio a statistical model has to be built that resemblances the phone of one person and a model for the phones of two persons. During the simulation of the data, thought has been given to generate a model of human calling behaviour, but many parameters can be thought of that are difficult to define which makes it problematic to translate the models to the real world.

3.10.5 Recommendations for further research

The methodology described is designed for the purpose of this thesis and are not known to have been performed before. First, it will obviously be very interesting to carry out the introduced methodology on real data to see if similar results can be found within actual data on mobile phone activity. Furthermore, it will be important to validate the developed statistical methods when different kinds of usage of two phones are taken into account and to study the effect of alternative measures of closeness.

With regard to discovering colocating phones, it could be informative to study time till dislocation in a large dataset. When phone records are available for many phones, the concept of time till dislocation could be further deepened. When phone records are available for a large number of people, it can be thoroughly investigated how likely it is that phones colocate for a longer period of time.

With regard to evaluating colocating phones, it could be exciting to design an experiment which could further validate the techniques that are designed

to find colocating phones. It would be interesting to select a person that uses two phones and see if the second phone can be found among all phone records using the same investigative techniques that are proposed in this thesis. One should then define patterns of movement of one phone, search for hits and follow those phones in time to determine which phones dislocate. Then, hopefully, the second phone will pop up among the hits and will be the only phone that shows colocation for a very long time. This experiment can also be performed using just one telephone, by separating its phone calls randomly, as if they are coming from two phones. That way a researcher does not have to search for a person that is carrying two phones and willing to cooperate.

When the phone records are available of two phones of which they suspect to be used by the same person, it is proposed to use permutation tests to generate 1000 phones that are in their use similar to the investigated phone. One could also think about a study design that allows to test not on artificial phone histories, but on actual existing pairs of phones. A second recommendation for further research is therefore to analyse a large number of random pairs of phones in order to study the closeness of their paths. When the phones are known to be *not* in the same hands it could be interesting to compare the closeness of their paths to a pair of phones that is known to be used by the same person. Perhaps it is even possible to take two phones into account that are not used by the same person but that are likely to colocate (for instance the phones of two people that live near each other and travel the same route to their work).

Overall it can be stated that several experiments can be thought of that could contribute to the question to what extent colocation analysis provides evidence of the identity of the users and to quantify the strength of this evidence. Additionally, the two performed projects could be further expanded since the described experiments and methods are freshly developed and open for improvements.

References

[1] (2010). R v T. EWCA Crim 2439. *http://www.bailii.org/ew/cases/EWCA/Crim/2010/2439.pdf.*

[2] Andersen, P.K. Borgan, Ø., Gill, R.D. & Keiding, N. (1993). *Statistical models for counting processes.* Springer.

[3] Berger, C. E. H., Buckleton, J., Champod, C., Evett, I. & Jackson, G. (2011). Evidence evaluation: A response to the court of appeal judgement in R v T. *Science and Justice*, 51: 43-49.

[4] Brodie, I. (2011). Thoughts and observations on prosecution use of cell site analysis. March 2011 Cell Site Blog. *http://www.cclgroupltd.com/Law-Enforcement-News/march-2011-cell-site-blog-attribution.html*, visited on 04/09/2013.

[5] de Montjoye, Y.-A., Hidalgo, C.A., Verleyse, M. & Blondel, V.D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports.* 3, 1376.

[6] Fenton, N. & Neil, M. (2012). On the limiting use of Bayes in presenting forensic evidence. *Draft.*

[7] Findley, K.A. & Scott, M.S. (2006) The multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, Vol. 2: 291-397, Univ. of Wisconsin Legal Studies Research Paper No. 1023.

[8] Finkelstein, M.O. & Fairley, W.B. (1970). A Bayesian approach to identification evidence. *Harvard Law Review, 83*(3), 489-517.

[9] Finkelstein, M.O. & Fairley, W.B. (1971). The continuing debate over mathematics in the law of evidence. *Harvard Law Review, 84*(8), 1801-1809.

[10] Lima, A., Domenico, M. De, Pejovic, V. & Musolesi, M. Exploiting regular data for disease containment and information campaigns strategies in country-wide epidemics. arXiv:1306.4534.

[11] Schneps, L. & Colmez, C. (2013). *Math on trial: How numbers get used and abused in the courtroom.* New York: Basic Books.

[12] Thompson, W.C. (2009). Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. *Law, Probability and Risk, 8*(3), 257-276.

[13] Tribe, L.H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review, 84*(6), 1329-1393.

[14] Tribe, L.H. (1971). A further critique of mathematical proof. *Harvard Law Review, 84*(8), 1810-1820.

[15] World Bank (2012). Information and communications for development 2012: Maximizing mobile. Washington, DC: World Bank. DOI: 10.1596/978-0-8213-8991-1.