Tampere University

Fryderyk Marek Pryjma

HOW TO OPTIMIZE AND CARRY OUT SPLIT ANALYSIS FROM LIVE GPS COVERED COMPETITIONS VIA THE TRACK-COURSE PLATFORM

ABSTRACT

Fryderyk Marek Pryjma
Bachelor's Thesis
Tampere University
Bachelor's Degree in science and Engineering
April 2021

The GPS coverage of sport competitions via the Trackcourse platform is currently limited by the lack of a live split generation feature. The study carried out aimed to discover the methods of allowing accurate split generation and optimalization of the processes occurring on Trackcourse's server while calculating splits.

The research followed an adjusted SPS (Structured-Pragmatic-Situational) approach to conducting case studies. During the study the extensive non-structured interviews were carried out along with a case observational study and the analysis of existing literature. The Trackcourse employees were interviewed, providing insights into current system design and previous approaches to the problem. The observational study was carried out by analysis of previous events carried out on the Trackcourse platform.

The research resulted in creation of three step procedure containing creation guidelines that shall be followed when designing the split generation feature for live GPS covered competition on Trackcourse platform. The steps identify the presence of a control point and ensure optimisation for the server's processes. These steps could be followed in creation of the live split generation feature also on other platforms.

The study was limited by enclosing it in the Trackcourse environment. Hence, it might not directly apply to other platforms. There is a possibility to apply the results in industry and use the feature in Trackcourse platform. There is also an opportunity for further study of the split calculation and estimation algorithms with the use of machine learning. The study can be also extended by creating sport specific control presence verification conditions.

Keywords: Trackcourse, GPS, Splits Feature, Sports

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

PREFACE

The Bachelor thesis has been a life changing journey. Even though the path has been bumpy there were lots of interesting views along the way. The guides along the way are those I would like to express my gratitude towards in that place.

In first place I would like to thank my Marko Helenius for being my course setter. Along with Marko there has been Laeticia Petit all along the way. They both supervised my thesis and process of its implementation. They were assisting me in times of doubt and in times when the direction seemed to be lost.

Moreover, I would like to thank Trackcourse representatives Jacek Morawski and Grzegorz Biolik who provided me with the topic for that thesis and have supplemented it with real life challenge and the journey associated with the search for its solution.

I would also like to thank Andrew Barnett who helped with language check for the Thesis. He has provided professional insights on the grammatical structure of the Thesis and the vocabulary used.

My gratitude, however, shall not end on that. There are multiple people that influenced that thesis along with creation of my self and reaching the stage of university career that allowed me for creation of that thesis. Among those that have added a rail to my rails I would like to highlight the influence of my parents.

To all those that have contributed to that thesis either via direct discussion or influence on myself throughout years I would like to thank thoroughly.

Warsaw, 3 March 2021

Fryderyk Marek Pryjma

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

SATNAV	Satellite Navigation
GNSS	Global Navigational Satellite System referring to all SATNAV systems and their augmentations
GPS	Global Positioning System
SPS	Structured–pragmatic–situational (SPS) approach to conducting case studies
GLONASS	Global Navigational Satellite System
TOA	Time of arrival
SMS	Short Message Service
MTB	Mountain Bike
ETA	Estimated time of arrival

1. INTRODUCTION

The sport industry is booming worldwide in 2020 [1], among the sports that are growing in popularity are triathlon, trail running and kayaking [2]. Those are endurance sports that are hard to follow live because of their often remote setting or the courses spanning across large distances. Hence, for following those races GPS technology is used.

Expanding the technology might bring additional value to spectators and better their possibilities to follow races. The GPS live coverage is a sector that can be further developed and there are multiple features that can be invented and developed. As for such motivation the study on split feature for Trackcourse platform has been initiated.

In the remaining part of the introduction presents the study aim, topic origin, brief info about the company, the research question and its core issues and the thesis layout.

1.1 Study aim

The aim of this thesis was to answer the most problematic current issue for Trackcourse [3]. The owners of the company [3] were contacted directly and enquired if there shall be any improvement they would like to have investigated. The company itself offers a GPS coverage for sport competitions. The Figure 1 represents the core product of Trackcourse.

Figure 1 Trackcourse [1] GPS coverage

1.2 Topic origin

The GPS coverage from sporting events have been present on the market for already some time. There were such competitions taking place in the Persian Gulf already in 1993 [4]. The competitions transmitted the boat race and offered augmented reality race coverage. Later on America's cup, the oldest trophy in sports [5], GPS coverage was provided online using the web service as it is done nowadays.

Nowadays the technology finds broad use in several disciplines of sports. There are certain limitations in the technology. For instance, it does not allow measuring rapid changes in velocity or very high-speed running [6], it cannot be used indoors either due to lack of direct contact with satellites. Nevertheless, it is a method broadly in use for team sports and it is considered valid for measuring the training load and tracking activity among players in team sports [7]. Besides those most popular sports that are team sports attractive on massive scale and growing to standards of national tradition and culture [8] the technology is present in earlier mentioned sailing [4][5][9], orienteering [10][11][12], trail running, triathlon[13] and other endurance sports allowing the excitement of following a live race via this manner [9].
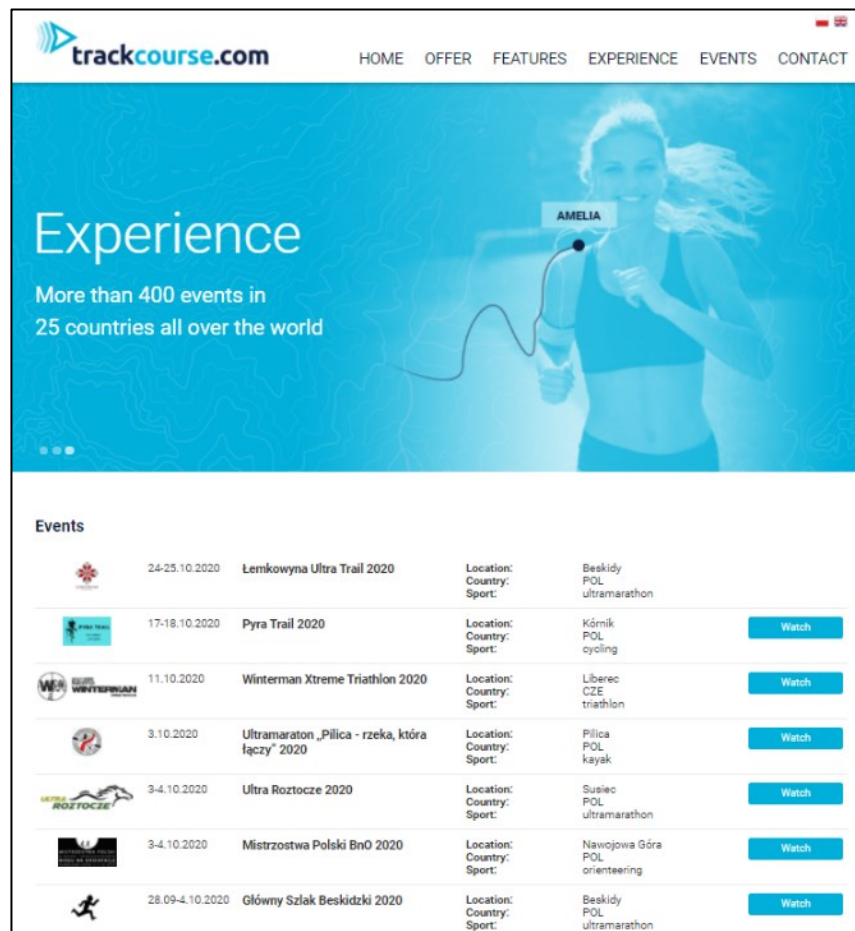
After observation of current industrial software the conclusion has been made that it tends to lack the live generated split times, the times on given point along the course allowing the spectator to see the current standing on given point. Therefore, Trackcourse

enquired on investigating their systems and the currently accessible technology to provide the best possible solution for the split analysis live system.

1.3   Trackcourse company

Trackcourse operates mostly on the Polish market. As presented in Figure 2 on the events tab one might see the events that the company has offered live tracking from. It has carried over 400 events in 25 countries around the world [3]. It offers live broadcast from multiple sports. Among those can be found: ultra running, kayaking, triathlon racing, cycling, orienteering, adventure racing and sailing. The company provides customers a solution being a combination of the software and hardware.

Trackcourse company focuses on several sports. However, in this study the scope was narrowed down to orienteering, trail running and cycling. The limitation has been made upon agreement with company representatives in order to exclude potential outer than the key factors. Factors specific to those three sports.

Figure 2 Trackcourse events [3]

The company is rapidly growing since it started working in 2014 and since then has had 1 598 810 users watching their events [3]. The company is Polish, and mainly operates on Polish, Czech and Baltic Countries markets. The features they offer in their products are illustrated in Figure 3.

Features

Universal web viewer
Follow the race live on any internet device - no additional downloads required

Sponsor logos exposure
Reserved spaces to display sponsor logos in viewer during your event

Replay feature
Available directly after the race, also with "mass start" option and performance analysis

Custom maps
Display the tracking on your own map or Open Street Map background (standard, terrain or sattelite view)

Social media
Event tracking website integration with most popular social media

GPX download
Instant GPX file download with each participants' track after the race

Handy event management
Easy and intuitive administration system developed in cooperation with race organizers

High-performance server
Operating on virtual cloud servers allows managing even the biggest events with thousands of live audience

Experienced staff
Full and professional support at all stages of your event

Solid trackers
Small, lightweight GPS units with great battery capacity providing up to 150 hours of tracking

Built-in memory
Position data collection by trackers even in areas with poor GSM coverage

Reliable GNSS chipset
GPS and GLONASS support connected with superior sensivity and fast time to first fix

Water resistant
Special waterproof bags allows tracking during triathlon, swimrun or sailing events

Multi-sport
Customized solution for any sport including triathlon, sailing, cycling, running, motorsports, etc.

Safety booster
Locate each participant in real-time in case of any unexpected situations

Figure 3 Features [3]

Their system itself relies on GPS (Global Positioning System) devices (trackers) that are illustrated on Figure 4. The tracker needs to be carried by participant of the competitions and their live track is displayed on the web application as illustrated on Figure 1.

Figure 4 GPS tracker [1]

1.4 Thesis's research question

The aim of the thesis was to answer the question of how to create splits feature that can be used during live covered competitions. The problem was that the so far suggested approaches were too heavy for the servers of the Trackcourse and the lag appearing due to their calculation has been too big. Therefore, the delay occurring has been preventing the spectators from following the competitions live but instead they often were presented with the results when they were already known from official time keeping system.

The goal of the research was to understand the current system and prepare a solution that could work without creating significant delay for the live covered races. The following goals were recognised at the beginning of the process: to minimize the processing time of the data flowing in the application, to minimize the load and number of queries the feature was creating in the system, to figure out an efficient method and algorithm for confirming that the user has visited a given control point.

The last goal was crucial to achieve since it assured that the results were reliable and trustworthy. The in-depth analysis of the goals and their nature will be in depth presented in Chapters 2 and 3.

1.5 Division to chapters

The thesis is divided into 4 chapters. Chapter 2 presents the theoretical background of the study. It will guide through the GPS technology, the process of creation of IT software and the way the client shall be treated and approached as well how the current system shall be accessed and the recommended feature environment investigated, in depth coverage of the use of GPS technology in sports with emphasis on the solutions provided by the case study company and the GPS software design comparison of different players on a market (for purpose of investigation whether the desired feature has been already created and how does it work).

The third chapter covers the research process. It starts from describing the investigation of current system as well as an interview carried with the employees and an in depth understanding of the desired feature as well as earlier attempts to solving the issue. Next it will guide the reader along the way the research has been carried with initial ideas for algorithm creation and verification on the basis of previous events whether the solution could work properly.

The last chapter highlights the limitations, findings, and topics open for discussion that resulted from the study and further study possibilities arising from that research.

2. BACKGROUND

The aim of this chapter is to discuss the technologies and theory behind the key concepts that are covered in the thesis. The chapter starts with an introduction to the technological innovation creation guidelines and choosing the approach for conducting the research process. Furthermore, the GPS technology is described, which is followed by the applicationof GPS technology to sport and the theory of sports that use it with emphasis on the sports serviced by the case company. Finally, the basic theory of interaction of the whole GPS system will be presented that will be later developed in the case study as well.

2.1 The technological innovation creation guidelines

For the research process and the implementation of the whole feature development process the study was based on the framework described in the article, regarding the SPS (structured-pragmatic-situational) study, by Pan Shan. L and Tan Barney [14] as well as on the extensive article recommending practises for software engineering case studies [15] by Runeson and Höst. The SPS method was chosen because of its case specific approach and reference to multiple previous studies, providing an essence of what they contained. This subchapter covers the methodology used in the study.

The study has been structured according to 6 steps described in the "Demystifying case research" article [14]. The steps are: Access negotiation, Conceptualizing the Phenomenon, Collecting & organizing the initial data, Constructing and Extending the Theoretical Lenses, Selective Coding and Writing the Case Report. The study [14], by default, offers 8 steps. However, in the case of the Trackcourse study the Confirming and Validating data and Ensuring Theory-Data-Model alignment steps are skipped because of the lack of previous studies regarding the live GPS coverage splits feature, which is required in those steps. Writing the Case Report does not have a special section in that study since it is obviously done by that thesis.

Each of the steps assures that optimal path for the research is followed. The steps structure the study at the same time as providing the flexibility needed for the case research since *"it studies contemporary phenomena in its natural context" [15].*

To prepare best for the first interviews, the extensive research on the Trackcourse company and the technology used in their platform has been done. The research

collecting and organizing the initial data mostly consisted of an interview of employees of Trackcourse and the survey of the current events and their application.

2.2 GPS technology

The following chapter is mostly based on a comprehensive summary of knowledge of the widely referred to GPS technology (referring to GNSS systems in general) and its alternatives „ Understanding GPS/GNSS: principles and applications" [16]. The aim of the section is to gather some essential data for understanding the matter of the problem.

The theory needed for understanding study may be summarised in the following sentences: there is an emitter and a receiver. A receiver (device used by a user) using radio signal connects with an emitter (satellite). Via the connection receives information that after processing provide it with its position. The receiver queries for the contact, and therefore acquires information of its position with specific time interval that is predefined. When the connection is lost there is no updated info about the position. When they reconnect the receiver might be already in another position and the track (set of positions) it covered from the previous query is lost.

2.2.1 GNSS

GNSS (Global Navigational Satellite System) is a general term referring to all satellite navigational systems and their augmentations [16]. Out of them, Global Positioning System (GPS) is the focus of this study.

GPS is the radio navigational satellite-based system (SATNAV) providing a user with navigational signals allowing him to determine factors such as velocity and position [16]. Besides widely known SATNAV system of GPS there is a Russian counterpart – GLONASS (Global Navigational Satellite System), the Chinese option BeiDou system and European Galileo.

2.2.2 Fundamentals of SATNAV

The GNSS use the TOA (time of arrival) to determine the position of the user. In that concept the measured variable is the time that was taken for the signal emitted by emitter (with known location) to reach the user. Then the signal propagation time (time interval) is multiplied by the speed of the source (i.e., speed of light). The measurement from different emitters enables to determine users' position. The position might be then saved with time and other factors on the device enabling broader use. The position is saved

using a reference coordinate system; there are multiple of those to be used for different applications.

2.2.3 Miscellaneous

In the usage of GPS technology there occur several errors and distractions. First of all, GPS requires direct contact between the emitter and receiver; blockages such as building roofs or ceilings, trees or even buildings might result in signal being absorbed or reflected which leads to loss of position info from the receiver. Dense vegetation might also result in similar issues.

Sometimes signal reflection might lead to one standing out from the others signal that will result in providing an odd position. These kind of deviations need to be taken into account when designing systems which process GPS data.

2.3 Sports with application of GPS technology

In the following section the basics of the sports which have been agreed for consideration (with Trackcourse representatives) in the research will be explained. Also, the key factors will be pinpointed which requirefurther attention for the creation of an algorithm.

Orienteering [17] is an endurance sport in which the orienteer must navigate using the map and compass while running. The map provides a competitor with information about the terrain: the height differences, the vegetation, the track network and many other things. Using the detailed map, a competitor must navigate through the controls to the finish. An example of an orienteering map is displayed in Figure 5.
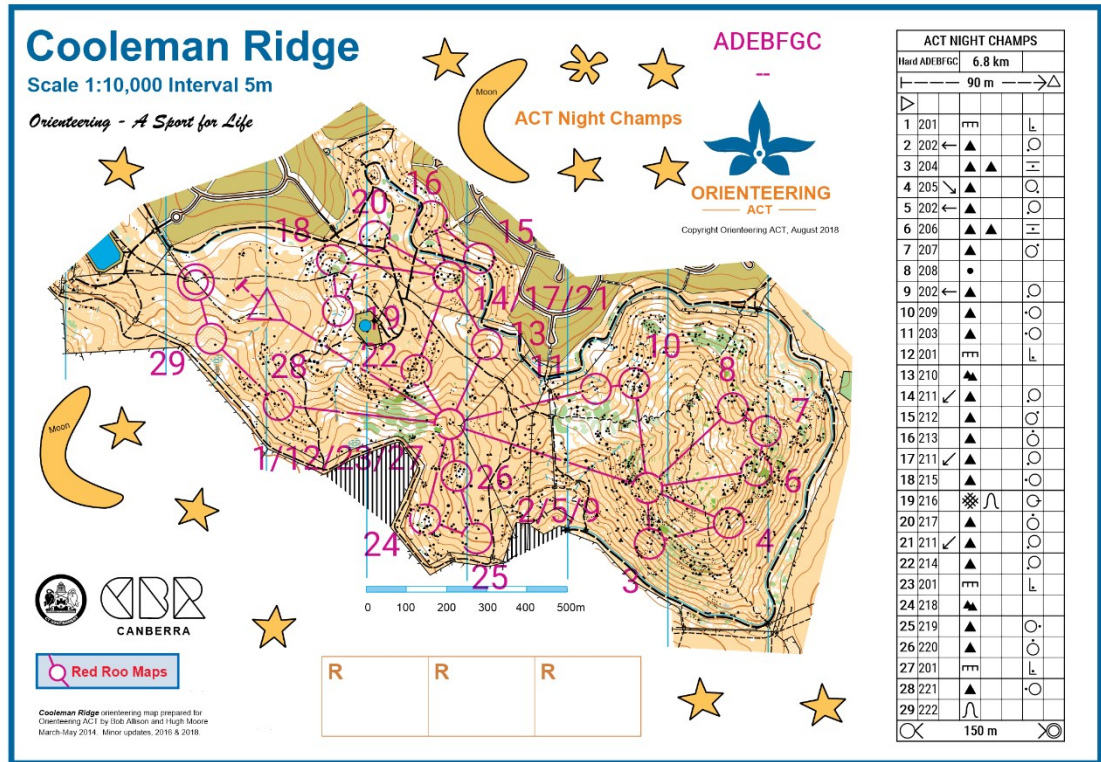
Figure 5 Orienteering Map

Trail running [18] is a running competition that takes place on different surfaces and locations. The races do not have specific limit of time or distance; however, the properly marked course must be easy for competitors to follow. In trail running use of hiking poles might be permitted.

Cycling [19] is the sport that can be done in either trail version – mtb (mountain biking) - or road version – road cycling. It is a sport that requires a use of bike for covering the given distance. Trackcourse events mostly cover ultra cycling events where there are cyclists operating on wide range of terrains.

2.4 The GPS software system design

The application providing tracking of sporting events works basing on a cloud application that can be accessed through the users' browser. The output in browser is the map illustrating the area of competition with tracking points representing the competitors as presented in Figure 1. The cloud-based application is processing all the data. It responds to queries from the user with information and updates the status of the race by receiving data from the trackers.

Trackers have their own memory and build-in SIM card. With that they send a text message with the data of the time and position in patches to the server. Server then

updates their location and their previous locations. The application works with delay being larger than the frequency with which the tracker updates the server with information. Most often it is 1 minute. This means that the situation seen on the users' screen displays the past state of current time adjusted by 1 minute.

The application therefore provides a visual representation on a map (from a bird's eye view) of the process of the race. That is an exciting insight to the race for the spectators.

3. CASE STUDY

This chapter follows the research methodology described in section 2.1 and so the 8 steps of SPS approach. The steps are in a way not linear to the thesis as a document since they refer to study process. The steps were the guide for the research process and a pointer towards the goal of the study.

3.1 Access negotiation

The study started from checking with the company owners of Trackcourse as to whether they had any challenges that could be the grounds for thesis.

The concept of a missing feature in their application and the lack of an idea for how to design it has been a leading phenomenon motivating the study. The feature was a live splits display from given points along the course of the competitions. Here a 'split' refer to the time at a given point along the racecourse (i.e., on 2km mark of 6km long running race). These points have specific coordinates, and they need to be visited in order. The access has been made through referral from influential benefactor [14], the owners.

3.2 Conceptualizing the Phenomenon

The core of the initial research is made up of non-technical literature (i.e. the internet articles)[14]. This was due to the lack of sources of scientific matter on the topic of live GPS sports event coverage. Also, the accessibility has been higher and the articles often referred more precisely to the topic.

At this stage, however, little information about the system is known, besides the fact that the illustrative data has been delivered in csv file format. That resulted in identifying best practices to handle data in csv files as one of first objectives. Another vital information has been the Python language requirement for providing the consistency in language operating on server to ease maintenance. Initially it was assumed that the data is saved on the server in csv format, which later turned out to be a mistake and a miscommunication.

The first leading problem coming from initial discussion with one of Trackcourse representatives, has been the csv processing time. In terms of handling this there has been Steve Hanly's [20] research carried out that displayed the potential of the panda library in the python language. The research has been similar in terms of amount of data

needed for processing. During the competitions carried out by Trackcourse there might be as much data processed as 20 million cells in csv files. Hanly's research offered insights on processing time of the csv files using standard methods and also illustrated potential gains from the use of the panda library. For further research, the article regarding reading large excel files in python has been beneficial [21].

Interesting insights for possible GPS processing methods were offered in the article "Review of GPS Travel Survey and GPS Data-Processing Methods." [22]. It provided comprehensive summary of available GPS data processing methods that might be used in later steps of SPS for construction of Theoretical Lens.

3.3   Collecting and organizing the initial data

After the initial research was done on the basis of the initial phenomenon introduction at the beginning of the study, the data gathering part started. The main conclusions were carried on the basis of a comprehensive two hours interview (notes from the interview can be found in Appendix A) with designers of the Trackcourse application and then proceeded with the study of previous events.

The interview enriched the study with crucial data concerning the current implementation of application, its technicalities, and the problems with earlier implementations of the split feature. The system design is illustrated in Figure 6. The GPS tracker saves its position on a local memory card every time interval 1. Time interval 1 is used for purpose of that thesis.The GPS tracker using cellular network sends an SMS to the Server with information on the number of entries from past SMS (entry here refers to position and time associated with that GPS ID). If there has been a break (lack of reception) in delivery to the server, then it first sends the newest entries, up to some not known by interviewees, number of last positions. The others (previous in time) will be sent in next SMS's according to company's protocols. The GPS tracker sends data in batches in time interval 2, that can be user specified.
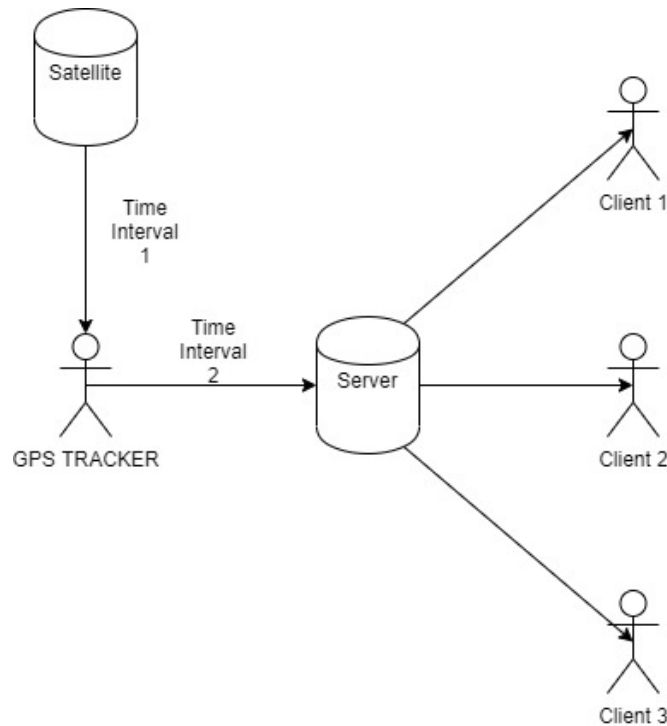
Figure 6 System illustration

During the interview it was also uncovered that the application relies on the Progress database. That has been a change to the initial interview result and an important detail for the theoretical lens creation. The csv file has been only the illustrative data patch for the initial recognition of the matter.

The clients, users wanting to follow online competitions, query the server for update on the position of the competitors. The whole system is browser based and therefore data is processed on the user's side. However, the queries load the system and as a result might lead to delays. That has happened in previous implementations of splits feature. The split system created too many additional queries overloading the system which therefore resulted in accumulating delay that on long competitions prevented the application from providing live coverage – too long delay for the service to be called live.

The observation study of competition has resulted in noticing the following errors/ system specifics. One of them is displayed on Figure 7. The dots with surnames and number refer to competitors. The black circle at the corner illustrates the control point (split measurement point). The triangle with number and surname represents a competitor whose tracker lost reception. It lost GPS signal on the picture 1. For illustrative purposes, the application depicts the path between last known position and the actual one. Therefore, on the picture 2 the triangle is on a straight-line way to position where it connected again, picture 3. This kind of signal loss needs to be taken into account when designing the practices to be followed while designing the split measuring feature.

Another was that in areas with tall building the signal tended to increase its error margin. Also the trackers sometimes sometimes happened to charge out. Besides this there was a case of a tracker falling out – remaining in the same position until the crew did not pick it up.

Figure 7 System failure

Moreover, the sports vary in terms of the speed with which the competitors might move (Figure 8). This impacts the distance a competitor might cover in a specific time interval. For illustration purposes 6 minutes has been used as time interval, but it could be any. Hence, the time interval 1, the one the GPS tracker records position of the competitor, must need to be taken into consideration – since, if too large, it might mean some corners could be skipped (like the one presented on Table 1 could be the result of too large time interval 1 too).

Table 1 Maximum speeds and distance covered

Sport	Orienteering	Trail Running	Cycling
Maximum speed (km/ h)	25	30	60
Distance covered in 6 minutes (km)	2.5	3	6

Another important factor observed has been accuracy of the devices. Based on the analysis of the previous competitions (around 10 full events were watched and observed for anomalies), it does not exceed 50 m from the firm point. The analysis was done by watching the orienteering competitions on Trackcourse. The control point on orienteering competition makes the competitor to come to the close proximity of it (< 1m). Hence it was a firm point for comparison of different positions provided from trackers. The time interval 1 was of 5 seconds on that competition and the average speed of fastest runners was slower than 6 min/km, which means the competitor's position could not be further than 6.7m away from the control point (5/60 (min) / 6 (min/km) / 2); in the equation the distance was divided by two since otherwise it would mean that the last position recorded

was on the control point and therefore the distance would be 0m. Therefore, to assume worst scenario the 56.7m mistake of GPS tracker needs to be taken into account for orienteering.

The battery of the devices has also been observed to have influence on the display. Occasionally the devices run out of battery. Then they stop emitting signals and stop on the display for the spectator. In these situations they are not capable of sending data to the server and therefore the split service is not able to detect the presence of that competitor on a control point.

Lastly important factors were noticed. The competitions strongly vary in length and number of competitors. There are also varying speeds. These factors impact the requirement of having fast appearing splits. In ultra running competitions, high refresh rate might not be essential.

3.4   Constructing and Extending the Theoretical Lenses

The research that has been carried on grounds of Trackcourse is innovative as it touches quite new ground of GPS races live tracking it has no previous theoretical lens that could be extended. Therefore, that study serves as a ground setter for the matter of split analysis approach on the races that have been earlier mentioned.

As an output of that research the procedures are created that might be later used for the software feature creation. Those are specifically focused on environment existing on the Trackcourse platforms and hence might not apply to the other implementations. The steps/ guidelines are listed in section 3.5 along with the motivations coming from the study for their existence.

3.4.1   Step 1 The conditions for confirming the presence at the control

First of all, there needs to be created conditions ensuring the competitors' presence at the control. Such presence needs to be ensured to confirm that the course has been completed according to the correct order and that no previous checkpoint has been skipped. Such situations have occurred in past as the famous example of cheating on Boston Marathon, cutting the course [23]. Hence condition one arises – presence on a previous control. However, there needs to be a back up option and therefore the system shall have an alert implemented as well – if by the end of expected time the competitor does not arrive to control. In that manner we approach condition two.

Condition two refers to both the expected arrival and maximum time competitors shall take to arrive at a control point. If maximum shall be crossed then red alert for the supervisor of the system shall be displayed. In this case the supervisor shall check if there have been GPS tracker malfunctions as the lack of presence may be caused by short lack of GPS tracker or satellite connection. The supervisor shall check the path of the competitor and if applied manually confirm presence on the control point. In this case the system shall refresh that competitor and continue checking for next competitors.

Expected arrival time is required, however, to limit the number of queries to the database storing information from current live covered GPS competitions. The feature shall not query for specific competitor unless his competing time goes over the minimum ETA. If the database does not allow for queries for specific competitors, then the minimum ETA for the whole competition shall be chosen.

Lastly the condition three arises. That is the most complex condition since it has a couple of steps that may need to be considered when designing a feature. Firstly, proximity needs to be taken into account. Generally, when designing that method there have been taken several scenarios taken into account. These are presented in Figure 8. The black crosses are the points where the GPS tracker records position, the green path is the path that competitor has chosen. The purple circle is the control point. Because of unpredictability of the courses on the events and all the different angles and courses that might appearing the simplest solution has been picked. The presence on a control shall be recorded for the competitor if his tracker records a proximity of

$$50[m] + \frac{(time\ interval\ 1\ [s])}{maximum\ average\ speed\ [min/km]}/2\ .$$

Maximum average speed is adjusted for the sport using i.e., data from Table 1. The following equation adjusts the error of the gps tracker and adds to it the extreme possible situation – where the control point has been reached exactly in between the two recordings of GPS tracker.
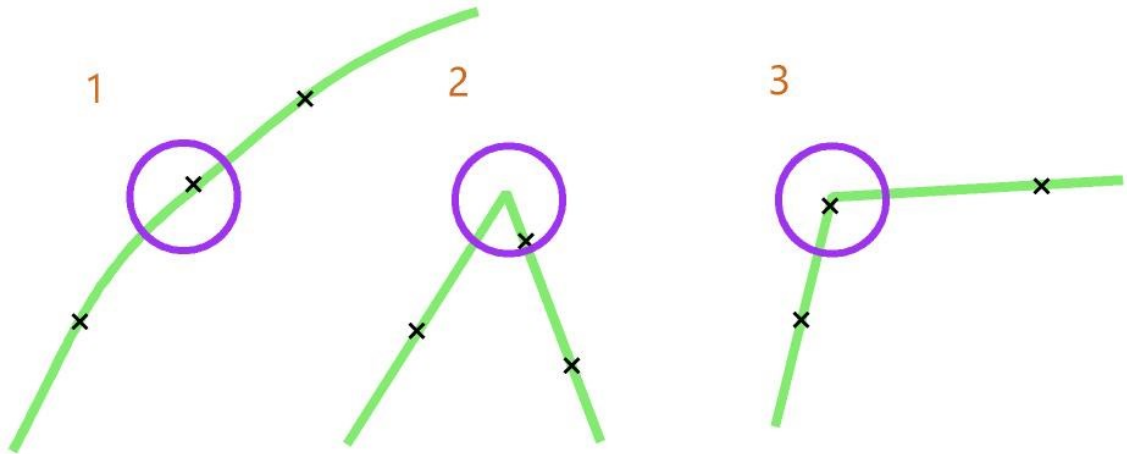
Figure 8 Possible tracks

That solution limits the effort of the supervisor of the event since it does not require him to add any extra control points for the verifying presence on a control. Other solutions that have been analysed require more human work and were more prompt for errors.

3.4.2   Step 2 Limiting the refresh rate of live results

Since the competitions might reach hundreds of competitors and might last for many hours they might result in big data. Then the constant requests to the database via this feature could lead to overload of the system and hence to delay in update of live results. That has happened in previous designs of the feature.

There are two ways to avoid that problem from happening. One is to extend time interval 1 to longer rate which will reduce number of recordings on the trackers. That will, however, impact the accuracy of the split service and the application's tracking ability in general. Therefore, the system shall have a predefined delay of 1 minute. Such delay would enable for all calculations to end by the time the next request to the server would be made and hence not overload the system. That would apply even in the most competitor loaded and longest competitions. The delay would also make it clear for viewers that the results will be updated within certain time intervals. One minute was chosen because of its compatibility with the system's design and also because of calculations provided in the research on MATLAB vs Python speed analysis [22].

3.4.3   Step 3 Ensuring maximum precision

To get the most accurate time measurement the feature could have an option for defining two extra control points (assisting ones). They shall be more less placed as the measurement points on image 1 of Figure 8 (black crosses outside control point along route). Then the average speed shall be estimated – straight line distance divided by

time covered between those points. That speed shall be used to adjust the time of central measurement – the straight line to the centre of control point could have been used for calculation of the time that would be taken to cover that route by competitor. Then the split time shall be adjusted by this calculation.

Moreover, the best tracking devices could be used that limit the 50m mark occurring in case of current system to 10m occurring in the best devices on the market. That however would increase the price of the system and could limit attractiveness to the wider public.

3.5   Selective Coding

The following section provides the data-theoretical lens alignment. The alignment is ilustrated in Table 2. The steps in first clomun are the results of the study discussed in subchapter 3.4. The last two columns present arguments to include such steps.

Table 2 Steps to be considered when implementing splits feature

Steps	Short description	Argument found in the interview	Argument found in the data analysis
Step 1 condition 1	Presence on previous control	Proximity would not be enough	Sometimes the path to the control leads through another
Step 1 condition 2	Minimum and maximum ETA	Reducing calculations	Sometimes GPS loses connection to satellite
Step 1 condition 3	Proximity	The main condition shall be simple	Wide variety of control approaching patterns
Step 2	Limiting refresh rate	Reducing traffic on the server's main database	Lack of negative impact on the following of the race
Step 3	Split time accurate estimation	Having the splits relatively accurate	Jump of recording points could be noted

4. CONCLUSIONS

4.1 Findings

The research aimed to discover the most effective way of making split analysis in the live GPS covered competitions using the data collected by GPS trackers. Table 2 is an output of the study providing collection of the most essential guidelines for developing a split feature for applications such as Trackcourse.

The most vital finding was the need of optimization. However, it was not supposed to be coming from the technology of the programming language but instead on extending the refresh rate from 0 to 1 minute; this was found to be the biggest need of impromevent in previous designs.

Moreover, an effective and simple 3 condition assembled control presence assurance has been formed. The more complex control presence condition systems would require more human work.

Those findings bring practical value to any company offering live GPS coverage . Following these guidelines they shall achieve a well functioning split generation feature.

4.2 Limitations of the study

Firstly, the limitation coming from the Trackcourse platform needs to be emphasised. It's design, the interviews and the survey of previous competitions have impacted the results of the study. The different solution' design might have resulted in other results because of implications coming from the technology use. Certainly, the findings bring value to other systems as well and might be used there for creation of the split feature. However, it is not certain whether they would be the most effective and accurate improvements.

Furthermore, because of high complexity of web application and server design no real application was developed, which was earlier considered as a possible goal of the project if the time would have allowed. There were only a few test runs working on the csv data sample. They, however, proved that this design has a great potential for successful implementation and delivery of end product.

Moreover, current design is limited by the the accuracy of GPS trackers. There is a possibility for going near the control and having the presence recorded. However, the later methods. i.e ensuring passing the control along the path, would not work in all the desired sports so the simplest one was chosen.

4.3   Further study

Most likely, machine learning could be implemented by control presence verification. This could work by analysing the competitors coming to the controls prior the competitor and patterns coming from their tracks. Hence, this could create more accurate conditions that would prevent cheating.

If several methods could be designed depending on sport of the event, then the sport specific methods could be created. I.e., for orienteering the proximity would need to be used, however, in trail running the additional condition of checking the direction of movement could be used to verify if the control has been visited and there has been no course cut made.

Lastly, the application could be designed according to the principles presented in this research. This could test whether the principles hold in practice.

REFERENCES

[1]   Why the Sports Industry is Booming in 2020 (and which key players are driving growth), Torrens University Australia, Available: https://www.tor-rens.edu.au/blog/why-sports-industry-is-booming-in-2020-which-key-players-driv-ing-growth#.YImgtNPivb0, Accessed on [28.04.2021]

[2]   New Report Provides Insights on What Sports are Growing, Gaining in Popular-ity, Sports Destination Management, Available: https://www.sportsdestina-tions.com/management/economics/new-report-provides-insights-what-sports-are-growi-15266, Accessed on [28.04.2021]

[3]   Trackcourse.com, Available: http://trackcourse.com/index.php?mnu=81, Ac-cessed on [22.03.2021]

[4]   Cooper, R. D., and D. Earl. "GPS-Augmented Live Television Coverage of Per-sian Gulf Boat Races." GPS WORLD 4 (1993): 40-40.

[5]   Sports fans worldwide will view America's cup racing live with virtual spectator technology. (2000, Feb 17). PR Newswire Retrieved from https://lib-proxy.tuni.fi/login?url=https://www-proquest-com.libproxy.tuni.fi/wire-feeds/sports-fans-worldwide-will-view-americas-cup/docview/449418935/se-2?ac-countid=14242

[6]   Rampinini, E et al. "Accuracy of GPS Devices for Measuring High-Intensity Run-ning in Field-Based Team Sports." International journal of sports medicine 36.1 (2015): 49–53. Web.

[7]   Theodoropoulos JS, Bettle J, Kosy JD. The use of GPS and inertial devices for player monitoring in team sports: A review of current and future applications. Or-thopedic Reviews. 2020;12(1). https://libproxy.tuni.fi/login?url=https://www-proquest-com.libproxy.tuni.fi/scholarly-journals/use-gps-inertial-devices-player-monitoring-team/docview/2437338459/se-2?accountid=14242. doi: http://dx.doi.org.libproxy.tuni.fi/10.4081/or.2020.7863.

[8]   Jozsa, Jr, Frank P. Global Sports: Cultures, Markets and Organizations, World Scientific Publishing Company, 2009. ProQuest Ebook Central, https://ebookcen-tral.proquest.com/lib/tampere/detail.action?docID=477168.

[9]   S. Honey and K. Milnes, "The annotated America's cup," in IEEE Spectrum, vol. 50, no. 9, pp. 36-42, September 2013, doi: 10.1109/MSPEC.2013.6587187.

[10]   Sirakov I, Belomazheva-Dimitrova S. Value îf technical trainings, their analysis ànd effects în òhe preparation process îf world elite orienteering competitors. Journal of Physical Education and Sport. 2018; 18:2127. https://lib-

proxy.tuni.fi/login?url=https://www-proquest-com.libproxy.tuni.fi/scholarly-jour-nals/value-îf-technical-trainings-their-analysis-ànd/docview/2160342885/se-2?ac-countid=14242. doi: http://dx.doi.org.libproxy.tuni.fi/10.7752/jpes.2018.s5321.

[11]   Norouzi, Monire. (2013). Application of GPS in Orienteering Competitions. Inter-national Journal of Mobile Network Communications & Telematics (IJMNCT). 3. 10.5121/ijmnct.2013.3402.

[12]   T. Zhihua, "The Application of GPS/GIS Navigation and Positioning System in Cross-Country Orienteering," 2008 International Conference on Computer Sci-ence and Software Engineering, Wuhan, China, 2008, pp. 582-584, doi: 10.1109/CSSE.2008.167.

[13]   Fister, Iztok & Fister, Dušan & Fong, Simon & Fister jr, Iztok. (2013). Widespread Mobile Devices in Applications for Real-time Drafting Detection in Triathlon. Inter-national Journal of Emerging Technologies in Web Intelligence. 5. 110-121. 10.4304/jetwi.5.3.310-321.

[14]   Pan, Shan L, and Barney Tan. "Demystifying Case Research: A Structured–pragmatic–situational (SPS) Approach to Conducting Case Studies." Information and organization 21.3 (2011): 161–176. Web.

[15]   Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering, 14(2), 131-164. doi:http://dx.doi.org.libproxy.tuni.fi/10.1007/s10664-008-9102-8

[16]   Kaplan, Elliott D., and Christopher J. Hegarty. Understanding GPS/GNSS : Prin-ciples and Applications . Third edition. Boston ;: Artech House, 2017. Print.

[17]   International Orienteering Federation webpage, [https://orienteering.sport/orient-eering/](https://orienteering.sport/orienteering/) [accessed on 03.04.2021]

[18]   World Athletics webpage, [https://www.worldathletics.org/disciplines/trail-run-ning/trail-running](https://www.worldathletics.org/disciplines/trail-running/trail-running) [accessed on 03.04.2021]

[19]   Global Cycling Union, https://www.uci.org/ [accessed on 03.04.2021]

[20]   Matlab vs python speed for vibration analysis, https://blog.endaq.com/matlab-vs-python-speed-for-vibration-analysis-free-download, [accessed on 06.04.2021]

[21]   Working with large excel files in pandas, https://realpython.com/working-with-large-excel-files-in-pandas/#reading-the-file, [accessed on 06.04.2021]

[22]   Shen, L., & Stopher, P. R. (2014). Review of GPS Travel Survey and GPS Data-Processing Methods. Transport Reviews, 34(3), 316–334. [https://doi-org.lib-proxy.tuni.fi/10.1080/01441647.2014.903530](https://doi-org.libproxy.tuni.fi/10.1080/01441647.2014.903530)

[23]   Murphy, D. (2019). Cheating the boston marathon is disrespectful to those run-ning an honest race: Bib mules and course cutters, beware.

APPENDICES

APPENDIX A. Main interview notes

1. Check the current implementations of the competitions and write down Your observations about different races in topic: triathlon, duathlon, sailing, cycling, running

2. Observations: sometimes the gps device looses signal – what happens there in the split system?, what happens if it finds itself back again, route is quite often linear so there is one exit and entry point to the split point – what about sailing?, what happens if there is the same couple of points – how the system should behave for such question?, what happens if there are so many competitors that there might be delay if checking current position for each of them at a time, what happens if somebody looses path, seldom there are any competitions requiring high accuracy , what if there are loops like in sailing one need to around two points – how to specify accurately the place of the bouy – gps?,

3. **(Trackcourse crew info)** If the gps gets lost it first sends actual position then the previous data. Information about app activation (procedure, whether it starts each time or is activated at the beginning) is up to me. The csv table is updated constantly during the competition or in patches. The gps id is different than the one on website I will get the needed data for work.

4. I need to ask the company about the software and the memory ram they have on server

5. **Ideas**
   a. Pandas for python for handling large csv files „*My testing showed the pandas.read_csv() function to be 20 times faster than numpy.genfromtxt(). And the genfromtxt() function is 3 times faster than the numpy.loadtxt(). It seems that you need pandas for large data sets.* " **Pandas**
   b. There seems not to be need to optimize the data more than it is with let say pandas or the row reader – one just need to check what's more efficient as the request for new data will be done let say every update time of the results – maximum time needed for processing the data so let say 20'' (needs to be checked and maybe could lineary by rising to the end of the competitions). But as long as with Pandas it will not take more than 20'' according to vibrations. **According to my calculations.**
   c. Seems like there is no concern about read time as long as it is not happening every some seconds **Emphasize it read out problems**

d. So every once in a while app should refresh the data on contrary to constant refresh the request to the system is more seldom **emphasize continue**

e. List comprehension to skip some lines of read in **not sure and have not yet checked if that lowers memory intake but I will try it during implementation**

f. My data set has 100 000 entries on 20 columns so it gives 2 milions of data points; I check the number of competitors on this race and time of it – 51 competitors and it lasted for around 18 hours (averaged). I checked some ultra races that can last longest there was one with 156 competitors and 30 hours of competition time. That should provide about 5 times more data points so time no longer then 5 second in setup of research conducted by… However need to test it in their environment compare and also know moreless the delay. Confirmed with company it might be up to 20 milions data points so about a million entries. **Are You aware of Your delay?**

g. Keep it easy the only logarithm You need is the distance from gps data point from the center of the circle after minimum time has passed from last control point. If that point is not known prior race the sufficient will be a gps device as a control point being the center of the zone or two devices being the line. There need to be some user predefined input concerning the specific of the competitions. The distance should be of 100m. Orienteering control circle is of 80m radius 100m provides sufficient accuracy here as well (checked with some competitions available by provider.). Provided we pick more distance controls than 200m from each other or with longer time difference between each other than 5'. **Logarithm described timer used here in python**

h. There shall be substitution of the current time (adjust if there appears closer data point within some specific time frame). **Is this ok? Or do You prefer to have only firm time updates.**

i. Times shall be updated every minute or data patch interval.

j. If the location does not fulfil the criteria (as in a minute You can surely surpass 200m there might be 2 options – one we increase the radius to the maximum speed they can move in the area multiplied by the patch data interval) we check the line created by two consecutive points – that is more demanding software wise as it demands having stored previous location somewhere. Assuming one would go 50km/h (on a bike) that would give a radius of 380m (not completely unrealistic providing the update of the results would be slowed down to 2.5' what equals to passing 380m with 8km/h in 2.5' speed and assures that the competitors have been checked for following time frame) **It might be wise to specify that the competitors will be visible on the result list once they passed the maximum time they should take from edge to the**

center + data sampling time to reassure that the time provided is the most accurate it can be.

    k. There need to be specified conditions and agreed for control point setting as maybe linear solution will make as possible to estimate the average time of passing the control point and then average the time it should take the person to be right in it – that will make as possible to estimate the accuracy of the solution. **Here i.e. possibility to upoad track or mark contemplatory points (entry and exit)**

    l. It will equal to only one request at a minute so it is ok, if company wants it less frequently might be changed manually so it is adjusted to competitions size interval of control points and the time of competition too. **What You want? That can be done with timeout function and will be adjusted to observed resulting time for data processign**

    m. Questions arising if I can update the file with new data minimizing the time it takes for read out of the data. How about choosing specific rows?

    n. Needs to be something done with the ability to download map as there should not be option. **Other point general time**

    o. **posgress database, api queries**

    p. Devices should be sorted according to start time on the app **other point**

    q. **Maybe You'd like to think of creating Your own celluar network? I was reading about long range weak celluar networks that could cover competitions with radius 10km to the transmitter – distributing them however along the course shall provide extreme coverage and one device price is like 100 euro**

    r. This kind of solution is used especially often infoor as well. They use indoor positionioning systmes to collect data.

    s. Gps is also used in sports so maybe upgrade to celluar network could allow to use of it in football/ cricket etc having the required range.

6. **Solutions to the problems discovered**
7. **Questions**
    a. Where the new files arrive to me it seems like to the beginning of csv file
8. **Meeting plan 16.11**
    a. First present findings

A. Time it takes to just read the data csv 0.0001271. || Copying just specific rows 0.9555217 ||

APPENDIX B. Supplementary interview notes

Notes from another discussion carried in November 2020.
TCPIP protocole
Parcer saves data on server being database with bufor
Database has data on all devices
bufor
when event finishes data goes from virtual database to local
saving each 5 seconds date hour minute secong; from the analysis of position change
there is a deviation;
we pick gps ids, app reads all positions for all on;
there is animation of movement between points;
displaying real data not processed from gps;
symphony;
parcer;
reduce number of entries;