# ABSTRACT

|  |
| --- |
| You must get the essence of what your report is about, usually in about 200 words. Most informative abstracts also have key parts in common. Each of these parts might consist of 1-2 sentences. The parts include: ♣ background ♣ aim or purpose of research ♣ method used ♣ findings/results ♣ conclusion |

The text of the abstract begins here. You may select this text and replace it with your own. The abstract counts as page iii, but as for all other pages with main headings (titles in all capital letters), a page number is not shown on it. The abstract is a required component of the thesis or dissertation, even in cases where individual chapters have their own abstracts. It should be a concise, carefully composed summary of the contents of the thesis or dissertation in which the problem is defined, the research method and design are described, and the results and conclusions are summarized.

There should be no more than 350 words in the abstract, which will normally make up about one and a half pages. No diagrams, citations, illustrations, or subheadings are included. If there are two pages in the abstract, the second page will be numbered iv, and this page number should appear centered at the bottom of the page.

# GLOSSARY

|  |  |
| --- | --- |
| SD | Succinct Data application |
| SMAC | Short Message Arithmetic Compressor |
| GUI | Graphical User Interface |
| XML | eXtensible Markup Language |
| XHTML | eXtensible HyperText Markup Language |
| IFRC | International Federation of the Red Cross |
| FRCS | Fiji Red Cross Society |
| NZ Red Cross | New Zealand Red Cross |
| iSMS | Interactive Short Message Service |
| UUID | Universal Unique IDentifier |
| SBD | Short Burst Data service (Iridium Constellation) |
| xsd | XML Schema Definition (used in the form tags) |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Introduction

|  |
| --- |
| This is the introduction part, the aim of this part is to introduce the Serval project, its background and main components, as well as the Succinct Data application, its background, its scope, how it works in a high level understanding. |

## Report structure

This report will go through my personal work on the Succinct Data application within the context of the Serval project. The structure is made in the way that every reader can find his way, depending on his knowledge on the project. As a matter of fact, it goes for each part of the report from the overview, motivations and background to the specific details.

The aim is firstly, through the remainder of the introduction chapter, to understand the motivations and the background behind the Serval Project and Succinct Data and develop a good overview. Then, the second chapter presents the specifications of the current Succinct Data application and the background and specification of the needed improvements which make the core of my work. The third chapter presents the design of the suggested solutions. The fourth chapter introduces the materials and methods used to implement the solutions. The fifth chapter goes into the specific implementation of each functionnality. The sixth chapter presents the results. Finally, the seventh and eighth chapter conclude the work and present some potential further improvements.

## The Serval Project

|  |
| --- |
| At the end of the part, the reader should have an overview of the Serval Project, knows the main idea, the why, and the main components of the implementation of the project (the how). |

### Overview and background

|  |
| --- |
| Explain generally in what consists the Serval project: What is the main idea? What is the aim? How it is generally implemented? What problem it is solving ? |

The Serval Project project has been launched in 2010 and consists essentially on direct communications between terminals (D2D) through their Wi-Fi interface. This technology finds its applications in several fields, for example:

In some third-world countries or countries under development, the network infrastructure is partly or totally absent (or accessible only by a minority of urban/higher class people). Then the cost to set up a totally new infrastructure would be higher than relying on direct communications (and satellite).

Natural disasters happen often in places where the network infrastructure is poor (ocean, islands, high mountains, etc.) or destroys the existing network infrastructure.

Some NGO as Red-Cross New Zealand have often assignments in places where the network infrastructure is poor and the reliability on direct communications between Red-Cross staff members and also the communication to the rest of the world through satellite would be critical.

### Key elements on the Serval Project

|  |
| --- |
| What are the main components of the implantation? How does everything work? Keywords : Mobile phones, Serval Mesh, Serval Mesh Extender, InReach devices, Satellites |

The Serval Project relies on many components to fulfill its requirements.

The first flagship element of the project is the Serval Mesh, an android application that is able to use the phone as a root, activates the Wi-Fi interface and communicate directly through this interface with other devices. The current possible types of communication are mainly voice calls through the VoMP (Voice over Mesh Protocol), text messaging with MeshMS (A service comparable to SMS), and file sharing.



Figure : An overview on the Serval Mesh interface

## Succinct Data

|  |
| --- |
| In what consists the Succinct Data application? What is the context and the why behind its implantation? How does it work generally? What is the idea of the compression? Results of the compression? |

The work of the current project is focused on Succinct Data, a mobile application that operates right after Magpi/ODK phone application in order to minimize the data produced by this application.

Magpi is a phone application that allows people to create surveys and polls that are materialized into a form with questions that can be a simple text field to fill or a multiple-choices question.

Succinct Data takes the XML file representation of the Magpi record (Which is also refer to as an “XML record”) and “extremely” compresses it to as less bits as possible, getting close to the Shannon entropy (i.e. the minimal and necessary quantity of information contained in the data).

### Overview on the Magpi application and the forms

|  |
| --- |
| What is generally the Magpi application? What types of forms can we create? How are the forms filled in? What are the files generated by the Magpi application for each action (create form, fill in a record …) ? |



Magpi is an application available on Android smartphones and on browsers to create surveys, polls and more generally to share structured useful data in a one-to-one, many-to-one or one-to-many way. Magpi have proven its utility in many fields including Education, Microfinance, Agriculture, Health and Disaster Relief. For example, iSMS (A service of interactive SMS) allows a speaker to share a poll with the audience on their phone via SMS and displays the results live. As well, DatAgro is a service in Chile that uses Magpi iSMS for a two-way and broadcast communication between local farmers and the agricultural cooperative, for sharing crop prices, weather updates or any other useful information.

Besides, Magpi surveys, which are claimed to be easy-to-set-up and low cost by their provider, have been used for example, in collaboration with the International Federation of the Red Cross (IFRC) to determine the cash distribution in an earthquake-ravaged region in Nepal in function of the needs. Finally, and that’s in what we’re focusing, Magpi is used by the members of the Fiji Red Cross Society (FRCS) and of the New Zealand Red Cross for real time (or at least extremely responsive) disaster relief in remote areas of the Pacific region.

Mainly, it is used by the first aid team to inform about the nature of the disaster, the number of casualties or injured people, and all useful data in order for the association headquarter to send the right number/amount of goods and rescue teams. The results of the poll are summed up in statistical graphics for easy reading and understanding.

The forms that are interesting us are written in XML language and interpreted by the Magpi application GUI. To be interpreted by the Magpi application, the data from a survey record always comes with 2 files:

* The form specification (cf. figure 2), an XHTML file for which the main role is to specify the form, the order of the elements, the type of each input and the possibilities for a multiple-choice question input (radio question).
* The form record (cf. figure 3), an XML file that contains *inter alia* the data content recorded by a user in the field, the timestamp of the record, and the record UUID.

../pictures/magpi/magpi_01.png

Figure 2: The form design and its corresponding XHTML specification file

The XHTML file remains the same for the whole life of the form (unless the design is changed) and is divided in many sections. For simplicity purpose I presented the specification for a form that would only contain the 2nd question (Type your name).

Firstly, we have the children of the <meta> tag (highlighted in blue) which specify the meta fields (i.e. general information about the form). Then, in the instance tag, the children of the <data> tag specify the the survey/form questions data. The <bind> tags contains attributes that informs us on the field, including the type (In our case it is a “xsd:string”). Then, the <xf:input> tags specify the question phrase and also the possibilities (In the case of a multiple choice question).

The fields that are important to understand are:

* The form ID (“formid”) which is a unique identifier for this form design.
* The ID of each field, often built from the question phrase (In our case type\_your\_name)
* The type attribute of the <bind> tags

../pictures/magpi/magpi_02.png

Figure 3 : The form record and its corresponding XML record file

The XML record is generated each time a form is filled with some data. Again for simplicity, I only present one field (but it’s the same for all fields). The XML record is much more simple than the XHTML because it only contains the actual recorded data while the XHTML specifies it. The record contains <meta> tag fields (in blue) which contain general information about the record (for example, the timestamp of the recording). The <data> tag fields contains the core information of the record.

Again, the most important fields to keep in mind are:

* The form ID which is the ID of the XHTML form specification linked to this file.
* The UUID which is the unique identifier of this record
* The ID of each question (the same that we can find in the form specification)

### Context and interests behind Succinct Data

|  |
| --- |
| In what consists the Succinct Data application? What is the context and the why behind its implantation? |

As seen in the previous Magpi overview, the Fiji FRCS and the New Zealand Red Cross first aid teams, and generally all disaster relief teams, use digital forms to report every important information in the disaster location (either by direct observation or polling the population). These forms are then sent to the NGO deciders so they can send as fast as possible the needed goods, materials and staff (this is sometimes a matter of life and death).

To communicate the important data to the deciders, field teams are using satellite communication devices such as *Delorme InReach*, or any other available mean (Cellular, SMS). There are two main reasons explaining the use of satellite communications and the need for a form extreme compressing application such as Succinct Data, one is practical and the other economic.

Practically, the reason is the same than the one behind the Serval Project. The area affected by the disaster is often hard to access, remote and has often a lack of telecommunications infrastructures or no infrastructures at all (destroyed or absent). The main constraint in disaster relief is the pace at which the teams, goods and materials are sent to heal injured people and save lives. Thus, we understand why the NGO such as the NZ Red Cross have no choice than using a time-guaranteed satellite service, such as the SBD service, to send the data.

Economically, the Short Burst Data service is provided by Iridium and is “a simple and efficient satellite network transport capability to transmit short data messages between field equipment and a centralized host computing system” [4]. And the fact is that this service is extremely expensive: In addition to the relatively cheap service fee of US$40,00, the data charge is billed US$ 1,09 per kilobyte [5]. Considering that the average size of a Magpi form is around 700 Bytes, this means that it would cost US$ 700 for 1000 forms, which exceeds the price of the smartphone and the satellite device combined (The used *InReach* device costs US$ 300), and thus is really expensive.

That is why the NZ Red Cross, which already uses the Serval Mesh for the communication between the team members in the field, requested for an application such as Succinct Data. Succinct Data compresses the form in an effective way, halving the size of the file to as little as 8.2% of its initial size [6].

The following figure explains the architecture in which Succinct Data performs its work, and the whole process of compression, from the field polling by the NGO field teams, to the receipt of the form by the NGO deciders.

../pictures/other/succinct_data_overview%20(1).png

Figure 4: Architecture and whole process involving Succinct Data [14].

### Overview on Succinct Data application

|  |
| --- |
| How does Succinct Data work generally? What is the idea of the compression? Results of the compression? |

Now that we know why Succinct Data was born and what is its main architecture, the network components involved and the used physical layer, we can focus on how it works at the Application layer.

As we know, SD takes XML records from Magpi (or any other forms provider such as ODK) and it compress it. But what is the global process besides the compression and what is exactly the nature of the compression? This is what this part will explain.

Actually, SD main objective is, as any other lossless compression system, to remove redundancy from the data to send (i.e. the XML record) at the sender side. The reversibility property of lossless compression will make the complete reconstruction of the data possible at the recipient side without any loss on the information. Assuming that both sides of the transmission have to keep this redundant information.

And it is the case, the only information actually unknown in the field is the value of each response in the form. All other information such as the order of the fields and the type of the fields can be completely implied. An analogy can be done with symmetric encryption: The compressed data is obtained by performing a series of reversible transformations that form the public key. Then, with the reverse operations, the other side reconstructs the data.

The following figure presents an overview on the operations made in the whole process:

../pictures/other/succinct_data_how_it_works.png

Figure : Overview of Succinct Data processes at the Application Layer

The following table explains the utility of each file involved in the Coding and Decoding Layer:

Table : Description of each file involved in the Coding and Decoding layer

|  |  |
| --- | --- |
| **File** | **Description** |
| Recipe | The recipe file is a shared file, obtained directly by extracting the information from the specification XHTML file, for which each line contains the following redundant information:  FieldName:Type:Minimum:Maximum:Precision,Select1,Select2,...,SelectN  “Select1, …, SelectN” is optional and corresponds to the N choices for multiple choices questions.  Combined with the information in the Stripped file, it allows the compression to a Succinct Data file. |
| Template | The template file is also a shared file and generated directly from the specification XHTML file.  It is a template for XML record files. It contains exactly the same information, except that the value of the field is replaced by “$field$”.  Thus, a field line looks like this: <field\_name>$field\_name</field\_name>  It is used at the very end of the decoding layer to reconstruct the XML record file from the Stripped Data |
| Stripped Data | The Stripped Data file is extracted directly from the XML record. It’s an intermediary file before complete compression and contains only key value pairs. It has the following content in each line: field = value.  It is used with the recipes as an input to the compression. The compression only keeps the salient information. |
| Succinct Data | The Succinct Data file is the binary file with the salient and lighter information to be sent via satellite or any other mean available (cellular, WiFi…). |

# General specifications

|  |
| --- |
| The specifications part enters with a lower level of understanding in the Succinct Data application. It explains what is the main architecture of the application, what are the components and the links between them (the android application, the jni library, xhcreate, strip, rexml, compress, decompress). Then it goes into the reason for what we need improvements and the specifications of this need (What is expected). |

## The current Succinct Data application

|  |
| --- |
| Part description : The current Succinct Data application explained in low level |

### The android application

|  |
| --- |
| Explain how the android application generally works, Inputs, Outputs, and the link with the compression application. |

### The compression application

|  |
| --- |
| Explain how the compression application generally works for each part explained in the overview in the part I (xhcreate, strip, compress, decompress, rexml). For each component, try to explain in in a nice way (If possible include graphics). The reader should already have an understanding of the purpose of each part of the work and now we feed him with the detailed way it work for each part. |

#### Creation of recipes and templates out of a form specification

#### Creation of the stripped file out of a record

#### Retrieving the record from a stripped file

#### Compressing a stripped file to a succinct data file

#### Decompressing a succinct data file to a stripped file

## Specification of the needs

|  |
| --- |
| There we go into the reasons for what we need improvements and the specifications of this need (What is expected). |

### Context and interests

|  |
| --- |
| The context behind the need |

SD users may have the need to use sub-forms in case of a survey that will be applied on an element that contains sub-elements. For example, in the case of a fire in a remote village containing 3 houses: A, B and C, the form will consist on general questions about the village (Village location, Village name, Number of households affected, …), then we can have sub-forms for each household in the village (House location, number of people damaged, list of materials damaged, …). The answers for a form will be the same for every sub-elements of the element (for example: The same answers for every house of the village). So, a solution for this issue that was implemented by Magpi is the sub-forms. That is why the support of sub-forms by the Succinct Data application is essential.

The SD application doesn’t manage the sub-forms. What we would like to do is to modify the code in the SD application so it is able to firstly recognize the tags and the elements of a sub-form and then process it correctly by transforming it into stripped data. Then, the code that transforms stripped data to binary data has to recognize the new stripped data structure and be able to transform it into binary data.

We want also that the method that creates a recipe from a form specification recognises the sub-form specification and calls itself recursively to create a recipe file for the sub-form.

### Functional specifications

|  |
| --- |
| Functional or technical specifications given for the solution : What are the functionalities to implement, the expected results |

# Design of the solution

|  |
| --- |
| In this part, we enter into the design explanation of the solution. After an introduction on the Serval project and on Succinct Data and the specifications of the needs, this part will explain the approach to implement the solution, the main algorithms, if possible state diagrams, any good graphical way to represent the design … |

## Overview

|  |
| --- |
| Overview of the solution : How will the solution be designed generally knowing that we have a legacy code and knowing how the legacy code works ? |

## Forms to recipes and templates

|  |
| --- |
| Part description : Design of the creation of recipes and templates from forms with subforms |

## Records to stripped data conversion

|  |
| --- |
| Part description : Design of the creation of stripped data from records with subforms and regeneration of xml file (reverse process). |

### XML record to stripped data

|  |
| --- |
| Part description : Design of the creation of strippeed data from records with subforms |

### Stripped data to XML record

|  |
| --- |
| Part description : Design of the regeneration of the original xml file from stripped data with subforms |

## Stripped data to succinct data conversion

|  |
| --- |
| Part description : Design of the compression of stripped data with subforms and reverse process (decompression) |

### Stripped data to SD

|  |
| --- |
| Part description: Compression |

### SD to stripped data

|  |
| --- |
| Part description: Decompression |

# Materials and methods

|  |
| --- |
| Description of the environment, the tools, and the methods for understanding the code, coding, testing. Also, it’s possible to give a planning of what was done from the beginning to the end |

# Implementation details

|  |
| --- |
| Part description: After presenting the design of the solution, we present here the implementation details for each part: What are the key variables and structures of data? How are they manipulated? What are the states and what is done for each state? … For each implementation part we can talk about some difficulties and how did we solve that. |

## Forms to recipes and templates

|  |
| --- |
| Part description : Recipe xhcreate |

## Records to stripped data conversion

|  |
| --- |
| Part description : Recipe strip and rexml |

### XML record to stripped data

|  |
| --- |
| Part description : Recipe strip |

### Stripped data to XML record

|  |
| --- |
| Part description : Recipe rexml |

## Stripped data to succinct data conversion

|  |
| --- |
| Part description : Recipe compress and decompress |

### Stripped data to SD

|  |
| --- |
| Part description : Recipe compress |

### SD to stripped data

|  |
| --- |
| Part description : Recipe decompress |

# Results

## Overview

|  |
| --- |
| Part description: Give an overview on the results and on what is achieved. What were the expectations, what did I achieve, what remains to do… |

## Test cases

|  |
| --- |
| Part description: Here give the test cases ran against the code to check its validity (either manual or automatic tests…). |

## Impacts / Discussion

|  |
| --- |
| Part description: What are the impacts of what I achieved |

# Conclusion

|  |
| --- |
| Part description: Talk about personal experience, path, general difficulties, achievements, sum up all the process and give a personal reflection about the work… |

# Further work

|  |
| --- |
| Part description: Further improvements |

# Bibliography

[1] P. Gardner-Stephen, A. Bettison, R. Challans, J. Hampton, J. Lakeman and C. Wallis. (2013) "Improving Compression of Short Messages", International Journal of Communications, Network and System Sciences, Vol. 6 No. 12, pp. 497-504.

[2] P. Gardner-Stephen, J. Lakeman, R. Challans, C. Wallis, A. Stulman and Y. Haddad. (2012) "MeshMS: Ad Hoc Data Transfer within Mesh Networks", International Journal of Communications, Network and System Sciences, vol. 5, no. 8, pp. 496-504.

[3] P. Gardner-Stephen et al. (2014) “Succinct Data: Extreme Compression for ODK Forms”, Global Humanitarian Technology Conference, pp. 70-77.

[4] P. Gardner-Stephen. (2011) “The Serval Project: Practical Wireless Ad-Hoc Mobile Telecommunications”

[5] P. Gardner-Stephen et al. (2013) “The serval mesh: A platform for resilient communications in disaster & crisis”, Global Humanitarian Technology Conference, pp. 162-166.

[6] P. Gardner-Stephen, R. Challans, A. Bettison., C. Wallis, J. Hampton and J. Lakeman. (2014) “Survey acquisition management tools for Android.” Internet: <https://github.com/servalproject/survey-acquisition->management [May 2016].

[7] P. Gardner-Stephen, R. Challans, A. Bettison., C. Wallis, J. Hampton and J. Lakeman (2014) “Short strings compression library” Internet: <https://github.com/servalproject/smac/tree/succinctdata>[May 2016].

[8] Magpi. “Advanced Mobile Data, Messaging, and Visualization anywhere on any device” Internet: <http://home.magpi.com/>[May 2016]

[9] P. Gardner-Stephen et al. “The Serval Project Wiki” Internet: <http://developer.servalproject.org/> [May 2016]

[10] P. Gardner-Stephen et al. “The Serval Project website” Internet: <http://www.servalproject.org/> [May 2016]

Other links/interesting :

<http://www.geekinsider.com/serval-mesh-chat-to-other-android-phones-without-network/>