Waveform Compression Algorithm

Eric Willie williee@oregonstate.edu
Joshua Wentzel wentzelj@oregonstate.edu
Bailey Sauter - sauterb@oregonstate.edu

Table of Contents

1. Design Impact Statement

1.1 Introduction

This document serves as an assessment of the design impact of the waveform compression algorithm that our group has designed for our project partner. The project consists of a software program that uses a variety of novel methods to compress a waveform from an oscilloscope to a smaller file size while maintaining certain signal characteristics. The following sections will assess the potential and actual impacts of this project with respect to five categories: public health and safety, cultural and social impacts, environmental impacts, economic factors, and any additional impacts.

This document may serve as a guideline for future considerations of people working on this project or another similar project. It may also be useful to the project partner to present considerations before the final implementation of the project.

1.2 Public Health, Safety, and Welfare Impacts

While remote oscilloscope signals may seem like a low-stakes situation for data compression, and therefore, not a public safety concern, there are situations where it could be a risk. Take the example of an oscilloscope monitoring the voltage waveform of a high-voltage converter. Small imperfections, noise, or spikes in that waveform may indicate imminent failure of the converter which may result in a safety hazard. However, if the lossy compression algorithm results in the loss of this warning, a safety risk may be posed to any workers or community members near the site being monitored [4]. Therefore, to design this algorithm ethically with respect to the public's welfare, the project team members must make explicitly clear to the client company the nature of lossy data compression and the risks that it poses. In doing so, the client company can implement the algorithm only in suitable and ethical situations.

A separate section will be devoted to environmental impacts, but the project team acknowledges that those impacts have a very real effect on public health and welfare.

1.3 Cultural and Social Impacts

There have been concerns raised about the morality of data compression, specifically lossy data compression, which is the style mentioned in this paper. There is an ethical principle of data that states that as much original content should be preserved as possible. This reduces the chance of corruption (and with it loss) and adds context or a big picture view to data that may be prone to misinterpretation [1]. Lossy data compression violates this principle directly as it is impossible to restore the original data in its complete form after lossy data compression.

This is especially relevant in a time where society and culture are filled with misleading and dishonest data and technology applications. The goal of this project is not to add to a societal distrust of technology and the people behind it. In order to do so, this project must be implemented in an honest and up-front way that acknowledges its shortcomings and potential for data loss.

1.4 Environmental Impacts

The main thing to consider when analyzing the environmental impact of data compression is energy consumption. There are two things to consider here: the potential for energy savings via compression (i.e. less battery life used, less bandwidth required to transfer data) and the overhead computational energy cost of running the compression/decompression algorithm. A study from a conference in 2003 [3] looked at this comparison and found that above a certain file size threshold, it would save energy (battery life specifically, in the study) to compress files. Based on this finding, the project partner would have some say in the environmental impact of this project. If they choose to blindly implement the algorithm, it may end up compressing some files that are small enough that the net energy consumption is increased by the algorithm's overhead computation. However, if implemented correctly, the algorithm has the potential to save energy and therefore, have a positive environmental impact on the world.

1.5 Economic Factors

The client for this project is a major player in the arbitrary waveform generation market. While not the specific application that this algorithm was designed for, a waveform compression algorithm such as this being developed for a company in the arbitrary waveform generation market could be used to greatly increase the speed, performance, size, and bandwidth capabilities of said company's products. With the aforementioned market projected to witness its highest-ever global growth by 2026 [2], the performance of this algorithm has the potential to offer the client company a leg up on the competition and greatly benefit its economic stakeholders. These stakeholders range from civilians with shares of the publicly traded company to executives and fellow employees of the company. This could create jobs for the company, resulting in a net positive economic impact in the community surrounding the client company.

1.6 Additional Impacts

There are a number of positive impacts that data compression as a concept can have on society. Decreased data size increases the amount of computing that can be done on a given hardware setup. This streamlining can lead to computational advances by making previously unreasonable algorithms possible. Assuming, these advances are in the right hands (a large and impossible assumption), this has the potential to lead to advances for society.

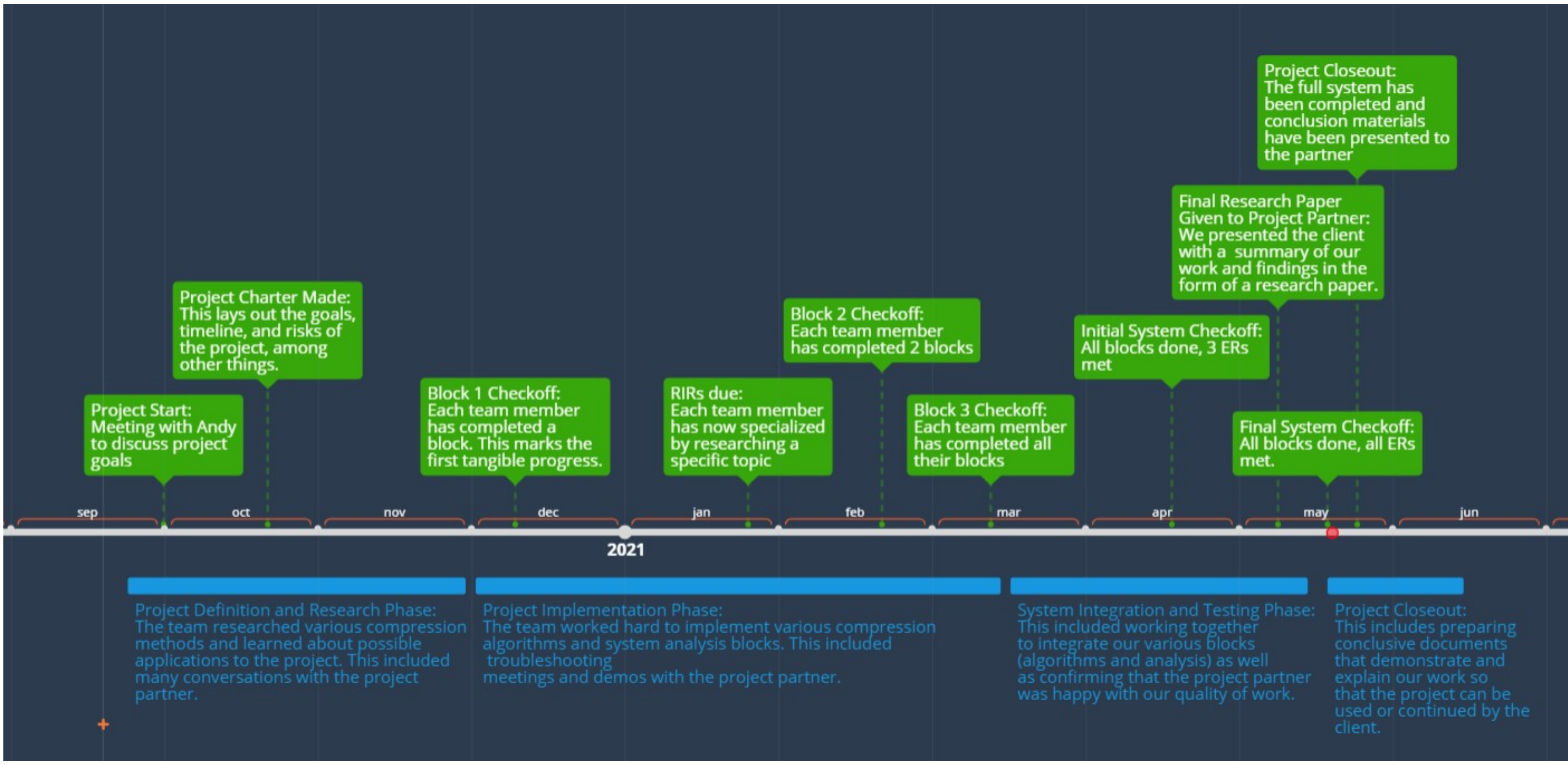
Therefore, the project team urges the project partner to be thoughtful about the sharing of any possible advances in data compression that stem from this project so that its impact is positive.

1.7 Conclusion

In summary, this project has many potential impacts. Despite seeming like a project that does not interface with the public, there are many ways it can impact a wide community. Therefore, the project team recommends that the project partner takes the necessary steps to ensure the final product is implemented in a fashion that makes those impacts positive rather than negative. This includes, among other things, making clear the risk of lossy compression, evaluating the most energy-efficient use of the algorithm, and presenting the resulting data in an honest and forward-facing manner.

2. Timeline

Below is a timeline of the major phases and milestone progress points over the course of this project.



3. Scope and Engineering Requirements

3.1 Scope

The purpose of our project is to research waveform compression algorithms and then, based on our research, either develop a new algorithm or optimize an existing one for use by automated remote oscilloscopes sending data to cloud servers. The goal of this project is to reduce the size of the data being transmitted in order to increase spending speed and reduce bandwidth.

This project is partially a research project in which we learned as much as we could about existing waveform compression algorithms. The second part of this project was using that new knowledge to develop our own algorithm that works well for Tektronix's intended use.

As stated earlier the final significant result for the stakeholders in the project should be increased speed and efficiency when transferring data from remote oscilloscopes for their work. The scope of this project is limited to the compression and reconstruction of the waveform. We will not be delving into the waveform measurement or the data transfer itself. Most of the project will be research and software development however, there is the possibility of using some hardware to generate the signal.

Some of the features within the scope of this project (reflected in the ERs) are compression ratio, time-domain waveform fidelity, algorithm complexity (time and space), frequency domain waveform fidelity, code modularity, and research results.

3.2 Engineering Requirements

Name	ER
1. Lossy compression ratio target.	The system will compress five test datasets by at least 75% of the original size for each dataset.
2. Compression Time Complexity	Any compression algorithm used must not have an exponential algorithm time complexity (i.e. Big O < O(2^n))
3. Decompression Time Complexity	Any decompression algorithm used must not have an exponential algorithm time complexity (i.e. Big O < O(2^n))
4. Time-domain waveform fidelity.	For five known test signals, their output decompressed version has the following characteristics: Waveform Mean, Min/Max, and STDEV must be within 10% of the original.
5. Frequency domain waveform fidelity.	For five known test signals, their output decompressed version has the following characteristics: The noise floor 1-sigma in the frequency domain must be within 10% of the original. The passband (up to Nyquist) amplitude should match within 10% (in linear scale).

6. Research paper.	A formal research paper will be written about the project including initial research, methodology, project code, results, and conclusion, and is signed off on by the project partner.
7. Algorithm Modularity	The system will define and utilize a protocol such that new compression/decompression algorithms can be incorporated.
8. Compression Space Complexity	The space complexity of all compression algorithms must not be larger than O(2^n).

4. Risk Register

ID	Description	Category	Proba bility	Impact	Performance Indicator	Respo nsible Party	Action Plan
R1	Group member gets COVID and can't work	Timeline	~5%	M	Group member health	Bailey	Retain
R2	The reconstructed waveform isn't suitable for Tektronix's use	Technical	~20%	H	Communication breakdown between the group and Tektronix	Bailey	Reduce
R3	Sample waveform for testing isn't realistic enough	Technical	~15%	H	Waveform sample has too few options to vary it	Bailey	Reduce
R4	Communication with project partner breaks down	Communic ation	~10%	H	Meetings and emails with project partners decrease in frequency.	Josh	Reduce
R5	One group member's computer isn't fast enough to run the algorithm	Technical	~1%	L	The algorithm doesn't work on one member's computer.	Josh	Avoid
R6	Our project doesn't meet the universal constraints	Technical	~50%	M	Universal constraints on the course website.	Josh	Avoid
R7	One group member is not pulling their weight	Timeline	~15%	M	Progress updates are not sufficient.	Eric	Reduce
R8	A mutually working group meeting time	Communic ation	~20%	H	Frequency of progress update meetings	Eric	Avoid

	can’t be found						

4.1 Lessons Learned Regarding the Risk Register

There were several things regarding the risk register that we did not foresee. The introduction of the vaccine greatly reduced the risk of a team member getting COVID. One thing we discovered was that we all use different versions of compilers for C++. This meant that some code only compiled on certain machines, adding an increased layer of complexity when integrating the project. This is a specific risk that should be considered moving forward with this project. We had some small communication and timing issues that were considered in the risk register but not completely avoided.

I would in the future start integration much sooner. This was probably the hardest part of the project and not a risk we considered at all. We spent a lot of time trying to integrate our pieces across compilers, languages, and computers and the end product was worse because we didn’t start sooner. Perhaps a better and more clear schedule accessible by the whole team could’ve eliminated this risk.

5. Recommendations

Reccomendation	Starting Point for Solutions
Only write code in one language. We ran into some problems where different team members were using libraries from specific languages. This meant that integrating the code was quite difficult because it wasn't in a uniform language.	Maintain a codebase on the cloud (ie Github, etc). This will allow team members and clients to stay up-to-date on the latest code and check that team members are using the same languages and protocols.
Compile code on a universal server. For example, not all C++ compilers are the same. This creates problems where parts of the code only run on certain machines.	Use a server that everyone can access that has one version of the compiler being used. This will increase ease of integration greatly.
Develop a graphical representation of compression results. Currently, the compression results are printed as text and provide statistics on compression and fidelity. A graphical representation would be easier to view at a glance and more impressive when presenting.	Python and MATLAB both have excellent graphical capabilities and will allow the presentation of the project results to be much-improved.
Test the compression algorithms on a wider range of test signals. We did not have a wide range of sample test signals from the client, which somewhat calls into question the thoroughness of our algorithm testing.	Look on the internet for sample oscilloscope waveforms and ask the client for more as well. Test algorithms on waveforms of different size (8-bit vs. 16 e.g.), type (sinusoidal vs. random e.g.), and format (signed vs. unsigned e.g.).
Ask the client for specific use-cases so that the compression can be better tailored. We designed our algorithms to work in a fairly broad range of uses, however, compression can be increased the more you know about a waveform.	Ask the client for a sample power user and tailor an algorithm to their case.
Examine more time-domain statistics. Our analysis offered only basic time-domain statistics. Providing more information in this realm would be more useful to the client.	Analyze the average difference between each sample in a signal. This will tell closeness better than a statistic like the mean.
Add a transmission protocol to replace the channel placeholder. This will better	Implement a basic network protocol like TCP and measure how fast different sizes of

simulate the real application and give info on if the compression is worth it or not.	waveforms can be transferred.
See how much compression can be achieved by ratcheting up the loss. This will push the limits of compression, something the client is very interested in seeing.	Forget about waveform fidelity for a moment and try to achieve maximum compression. Then see what results that has on the reconstructed waveform.

References

[1] International Association of Sound and Audiovisual Archives. (2020). *Data compression and data reduction.* [online] Available at: https://www.iasa-web.org/tc03/11-data-reduction [Accessed 18 Jan 2021]

[2] Data Bridge Market Research. (2021). *Arbitrary Waveform Generator Market to Witness The Highest Growth Globally In Coming Years.* [online] Available at: https://ksusentinel.com/2021/01/19/arbitrary-waveform-generator-market-to-witness-the-highest-growth-globally-in-coming-years-2026-tektronix-inc-teledyne-lecroy-bk-precision-corporation/ [Accessed 18 Jan 2021]

[3] Rong Xu, Zhiyuan Li, Cheng Wang and Peifeng Ni, "*Impact of data compression on energy consumption of wireless-networked handheld devices,*" 23rd International Conference on Distributed Computing Systems, 2003. Proceedings., Providence, RI, USA, 2003, pp. 302-311, doi: 10.1109/ICDCS.2003.1203479.

[4] NTT Inc, *Understanding the Dangers of Working Near High Voltage.* [online] Available at: https://www.nttinc.com/blog/understanding-the-dangers-of-working-near-high-voltage/