

---

# PKZip FILE PARSER

PA193 TERM PROJECT - STAGE I

FATJONA PODA

MARCO GASPARINI

MICHAL RAJČAN

# Pkzip File Format Overview

Local file header 1
File data 1
Data descriptor 1
Local file header 2
File data 2
Data descriptor 2
...
Local file header n
File data n
Data descriptor n
Archive decryption header
Archive extra data record
Central directory

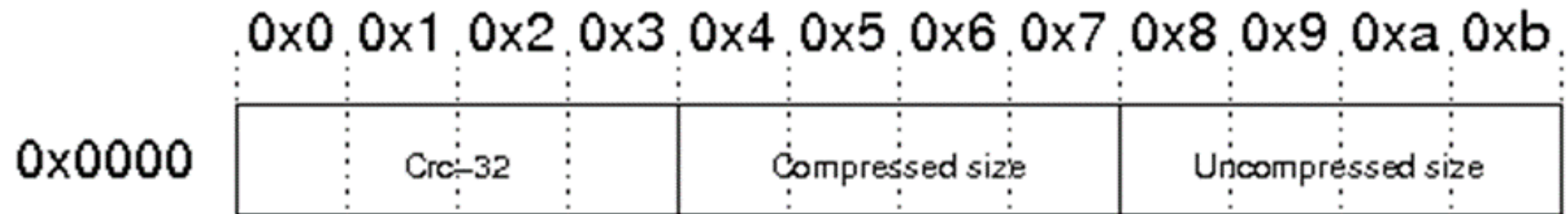
## ASSUMPTIONS AND LIMITATIONS (1/2)

- The architecture on which the parser is compiled is little-endian.
- The file does not contain a Archive Decryption Header.
- The file does not contain an Archive Extra Data Record.
- The Zip64 format is not supported.

## Local File Header Structure

	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7	0x8	0x9	0xa	0xb	0xc	0xd	0xe	0xf
0x0000	Signature				Version		Flags		Compression		Mod:time		Mod: date		Crc-32	
0x0010	Crc-32		Compressed size				Uncompressed size				File name len		Extra field len			
0x0020									File name (variable size)							
0x0030									Extra field (variable size)							

# Data Descriptor Structure



# Central Directory Structure

	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7	0x8	0x9	0xa	0xb	0xc	0xd	0xe	0xf
0x0000	Signature				Version	Vers. needed		Flags		Compression		Mod.time		Mod.date		
0x0010	Crc-32				Compressed size				Uncompressed size				File name len		Extra field len	
0x0020	File comm. len		Disk # start		Internal attr.		External attr.				Offset of local header					
0x0030	File name (variable)															
0x0040	Extra field (variable)															
0x0050	File comment (variable)															

# End Of Central Directory Record Structure

	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7	0x8	0x9	0xa	0xb	0xc	0xd	0xe	0xf
0x0000	Signature				Disk number		Disk # w/cd		Disk entries		Total entries		Central directory size			
0x0010	Offset of cd wrt starting disk				Comment len		ZIP file comment (variable)									

## ASSUMPTIONS AND LIMITATIONS (2/2)

- The file does not contain data descriptor fields
- There are no integrity checks performed for:
  - CRC
  - consistency between LFH and CD fields

# STRUCTURE

Based on functionality, the code is divided in ten different files:

- Makefile
- headers.h
- error\_handling.h
- parser.h
- printing.h
- main.c
- headers.c
- error\_handling.c
- parser.c
- printing.c

# DESIGN

- `uint*_t` instead of `char*`
- Functions `sure_fread` and `sure_malloc` are used to handle errors
- The code is compatible with both ,Windows and Unix Systems

# CONTRIBUTION

- All the team members contributed on the project.



THANK YOU!

