Recovery of ZIP Archive Files

```
Offset      0  1  2  3  4  5  6  7   8  9  A  B  C  D  E  F
00000000   50 4B 03 04 14 00 00 00  08 00 21 77 68 48 34 36   PK        !whH46
00000010   7F 14 7F 80 00 00 50 C3  00 00 0C 00 00 00 46 69      €  PÃ      Fi
00000020   6C 65 20 28 34 29 2E 74  78 74 1C 9D 09 9A AD 2A   le (4).txt   š-*
00000030   B2 85 47 46 DF 37 8A 20  9D C0 FC C7 91 B1 F3 D6   ²...GFß7Š  ÀüÇ'±óÖ
00000040   57 EF D5 3D 27 73 6F 85  88 B5 FE A5 88 3A 9C DE   WïÕ='so...ˆµþ¥ˆ:œÞ
```

Introduction

ZIP is one of the most widely used compressed file formats. It is universally used to aggregate, compress, and encrypt files into a single interoperable container. We have developed a methodology for recovery which has been embedded into an Intelli-Carve® recovery profile. Our software has the ability to read and validate ZIP archives directly from a stream.

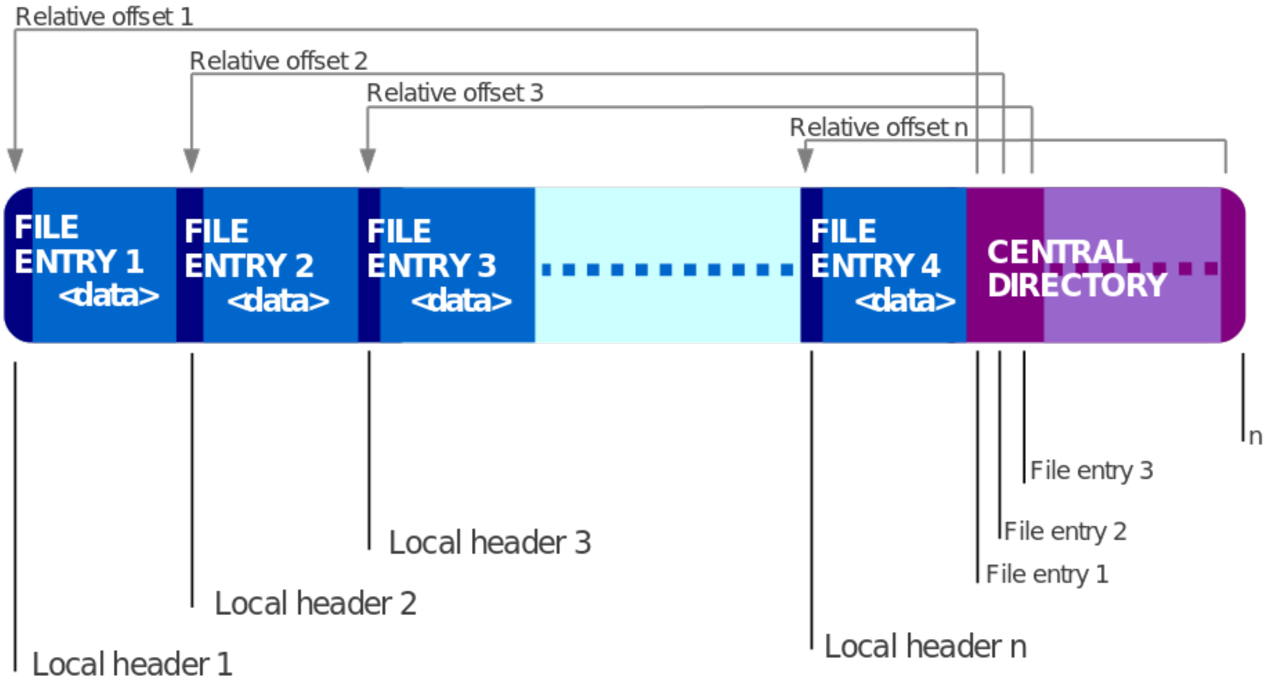
In addition to being used as a compression file format, ZIP is also used in a number of proprietary file formats such as those used for the following file types:

- Microsoft Word from 2007
- Microsoft Powerpoint from 2007
- Microsoft Excel from 2007
- OpenOffice Documents
- StarOffice Documents
- Adobe AIR installation packages

Data Structures

ZIP files contain three main objects:

- End of Central Directory structure
- Central Directory comprising of one or more Central Directory File Header structures
- One or more Local File Header structures



File Recovery

Our Blade® Intelli-Carve® profile understands the structure of the ZIP Archive and can load the data directly from a stream. A verification process checks that the End of Central Directory structure contains valid data; if this verification process is successful, Blade® then reads the Central Directory File Headers. On completion, the Local File Header structures are read and verified. During this process, Blade® decompresses each stream to ensure there are no errors. Once the data has been verified, we identify the file type by examining the data contained in some of the compressed streams. The file is then written out with the appropriate file extension.