

SMPTE STANDARD**Unique Digital Media Identifier (C4 ID)****Table of Contents**

1	Scope	3
2	Conformance Notation	3
3	Normative References	3
4	Terms and Definitions	3
4.1	C4 ID	4
4.2	Secure Hash Algorithm	4
5	C4 ID Specifications	4
5.1	C4 Base58	4
5.2	C4 Digest	4
5.3	C4 ID	4
5.4	Calculating a C4 ID for Non-contiguous Blocks of Data	5
Annex A	Pseudocode to Compute a C4 ID and a C4 Digest (informative)	6
A.1	Compute C4 ID from C4 Digest	6
A.2	Compute C4 Digest from C4 ID	6
Annex B	Diagrams of C4 IDs for Non-contiguous Blocks of Data (Informative)	7
	Bibliography (informative)	9

Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in its Standards Operations Manual. This SMPTE Engineering Document was prepared by Technology Committee 30MR.

Intellectual Property

At the time of publication, no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

C4 IDs are unambiguous, universally unique and consistent identifiers for files or units of data. Two organizations will compute exactly the same C4 IDs for identical data. C4 IDs have no need for a central registry or other external information.

C4 IDs are derived directly from the digital material they identify, so, unlike assigned IDs that are linked to the identified object by an external agency, C4 IDs are perpetually and unambiguously linked to the identified object and can always be regenerated with absolute certainty at any time.

This standard describes the C4 Base58 character set, the format and construction of the C4 ID for a single contiguous unit of data (e.g. a file) and method of construction of a C4 ID for non-contiguous units of data (e.g. some or all of the files in a folder or all or some of the data in a package).

C4 IDs for non-contiguous units of data are especially useful when archiving or transferring data because data verification can be done folder-by-folder instead of file-by-file. A single change of one bit in a sub-folder many levels below a folder having a C4 ID of this type would cause a change in the value of the highest level C4 ID. So if the C4 ID is the same for a folder in one place as a folder in another place, it indicates the data in the folder and all of its children have the same data.

At the highest level, C4 IDs can also assure all of the applications and other files beyond film or video data match.

C4 ID can be used to reduce the transmission costs of data across a network. It can do this in at least three important ways:

1. By keeping C4 ID values for all relevant files on remote and local systems, re-transmission of duplicate files can be avoided.
2. By confirming C4 ID values after transmission, systems can detect transmission errors.
3. By computing the C4 ID for a file immediately upon creation, sites can compare the results of deterministic processes without transferring the resulting files. For example, unpacking an archive file (or restoring a backup) can be validated by comparing C4 ID values of the resulting files with those computed at the time the archive was created.

NB: C4 is an abbreviation of Cinema Content Creation Cloud, reflecting its origin in cinema creation. The application of C4 ID is generic to all digital media files and is not constrained to cinema files. Because it is self-identifying, and for compatibility with already existing implementations, the prefix "c4" is maintained in the standard formulation, as is the "C4 ID" nickname in reference to that formulation.

1 Scope

This standard describes the C4 Base58 character set, the format and construction of the C4 ID for a single contiguous unit of data (e.g. a file) and method of construction of a C4 ID for non-contiguous units of data (e.g. collection of files or blocks in a stream).

2 Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword "reserved" indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword "forbidden" indicates "reserved" and in addition indicates that the provision will never be defined in the future.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; then formal languages; then figures; and then any other language forms.

3 Normative References

The following standards contain provisions, which, through reference in this text, constitute provisions of this engineering document. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this engineering document are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

ISO/IEC 8859-1:1998, Information Technology – 8-Bit Single-Byte Coded Graphic Character Sets – Part 1: Latin Alphabet No. 1

ISO/IEC 10118-3:2004 Information technology – Security techniques – Hash-functions – Part 3: Dedicated hash-functions

4 Terms and Definitions

For the purposes of this document, the following terms and definitions apply.