Secure Content Identification Mechanism Requirements

Candidate Version 1.0 – 16 Dec 2008

Open Mobile Alliance

OMA-RD-SCIDM-V1_0-20081216-C

Contents

Figures

Tables

1. Scope (Informative)

This Requirement Document (RD) contains use cases and defines the requirements for the Secure Content Identification Mechanism enabler to identify all kinds of content including both premier and user generated content, for various applications such as copyright verification, software authentication, and content filtering, etc. The following areas will be covered in this RD:

- Content identity registration/query/verification/certification/error reporting
- Content ID assignment
- Secure content identification/authentication mechanisms

Secure Content Identification Mechanism enabler will reuse as much as possible existing technologies. Some requirements may be shared between multiple OMA WGs, where extensions to other enablers may be needed.

2. References

2.1 Normative References

[RFC2119] “Key words for use in RFCs to Indicate Requirement Levels”, S. Bradner, March 1997, URL:http://www.ietf.org/rfc/rfc2119.txt

2.2 Informative References

[OMADICT] “Dictionary for OMA Specifications”, Version x.y, Open Mobile Alliance™, OMA-ORG-Dictionary-Vx_y, URL:http://www.openmobilealliance.org/

[OMA-CBCS-RD] “Categorization Based Content Screening 1.0”, Open Mobile Alliance™, OMA-RD-CBCS-V1_0, URL:http://www.openmobilealliance.org/

[OMA-MobAd-RD] “Mobile Advertising Requirements”, Open Mobile Alliance™, OMA-RD-Mobile_Advertising-V1_0, URL:http://www.openmobilealliance.org/

3. Terminology and Conventions

3.1 Conventions

The key words “MUST”, “MUST NOT”, “REQUIRED”, “SHALL”, “SHALL NOT”, “SHOULD”, “SHOULD NOT”, “RECOMMENDED”, “MAY”, and “OPTIONAL” in this document are to be interpreted as described in [RFC2119].

All sections and appendixes, except “Scope” and “Introduction”, are normative, unless they are explicitly indicated to be informative.

This is an informative document, which is not intended to provide testable requirements to implementations.

3.2 Definitions

Content Metadata	Information about a content item, such as content attributes (e.g. the title, ID, associated category, description, or perspectives) or data associated with the content (e.g. geo-information about where the content was produced).
Content Fingerprint	A short “summary” derived from the content that can uniquely identify the content in most contexts.
Content Fingerprint Extractor	Entity that extracts a Content Fingerprint from the content that can uniquely identify the content.
Content ID	A symbol (e.g. number or string) that establishes the identity of the content of the services to be used during its lifecycle (e.g. the assignment, the registration, the query, the verification, etc.).
Content ID Certificate	A certificate used to verify the identity/integrity of the content.
Content Identity Manager	An entity that manages content registration, responds to content identity query and verification request, and issues Content ID Certificates.
Content Provider	Entity that provides content for user consumption, usually in exchange for profit. This includes traditional content providers such as label companies, as well as individuals.
Digital Watermark	Auxiliary data that is imperceptibly and persistently embedded into an original content such as image, video and audio. This auxiliary data can subsequently be recovered from the watermarked content. Digital Watermark can be used to identify a content item, to verify its integrity, to authenticate the content with embedded copyright mark, to include meta data, etc.
SCIDM Client	An entity that makes requests to the CIM for content registration, content identity query and content verification.

3.3 Abbreviations

CIM	Content Identity Manager
CME	Content Management Entity
ID	Identity
OMA	Open Mobile Alliance
RD	Requirement Document
SCIDM	Secure Content IDentification Mechanism

UGC	User Generated Content

4. Introduction (Informative)

Today, mobile content spreads all over the mobile service world. How to securely and efficiently identify a mobile digital content is becoming a more and more important issue, and is expected to have potential impact on the successful deployment of mobile services. Secure content identification makes managing intellectual property in a networked environment much easier, and allows the construction of automated services and transactions. With the recent development of Web2.0, secure identification of user generated content becomes an important concern as well. The potential applications of secure content identification include charging, content search/management, automatic content monitoring for copyright verification and usage statistics, content filtering/blocking, content tracing, selective recording/playback, remote triggering of ads in broadcast chains, etc.

Secure identification and authentication of digital content would allow secure content transactions between all entities (e.g. Content Provider, content distributor, service provider, operator, enabler, end user) in the service environment, resulting in a more trustworthy and efficient service/transaction environment. This will greatly benefit all parties involved.

This Requirement Document (RD) defines the requirements for the Secure Content Identification Mechanism enabler to provide content identification service to various applications.

4.1 Content Life Cycle

SCIDM aims to provide secure content identification service to facilitate the management of content for various applications, such as copyright protection, infringement content filtering, etc. To be securely identifiable, content should be registered to a Content Identity Manager (CIM). A simple content life cycle is described here that includes the following states:

a) Created

b) Published

c) Registered

d) Identified

e) Repurposed

f) Managed.

Note that states and actions in a general content life cycle that are not closely related to content identification are not included here. The following figure shows the state transition diagram, which includes the states (the ellipses) and actions (the arrows between ellipses).

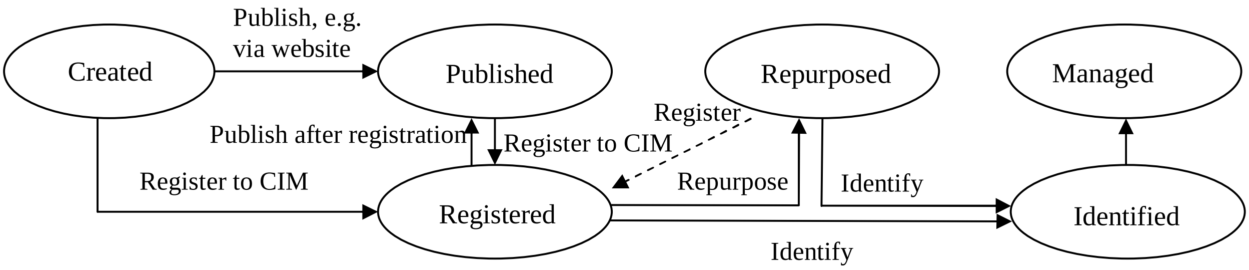

Figure 1: A content life cycle diagram

After a content item is created, it will be published, e.g., via a content hosting site. To be securely identifiable, content together with some associated metadata should be registered to the CIM. The registration can be done before or after the content has been published. Upon registration, a CIM-issued certificate that binds the content with some of its metadata can be returned to be associated with the content. After registration, the content can be securely identified by querying the CIM, or by directly verifying the CIM-issued certificate that is attached to the content. The registered content may also be repurposed/modified through, e.g., compression, trans-coding, reformatting, etc. The repurposed/modified content, as their original has been registered, may also be securely identifiable. The repurposed/modified content can optionally (represented by the dashed line in Fig. 1) be registered with the CIM again with a new ID. Once the content is securely identified, it can facilitate more effective and efficient content management (e.g., copyright control, filtering, searching, etc.). Note that the "Managed" state represents a state in which a specific content management operation has been applied to the *identified* content.

4.2 SCIDM Enabler Ecosystem

The following diagram shows the main actors and their roles in the SCIDM enabler ecosystem.

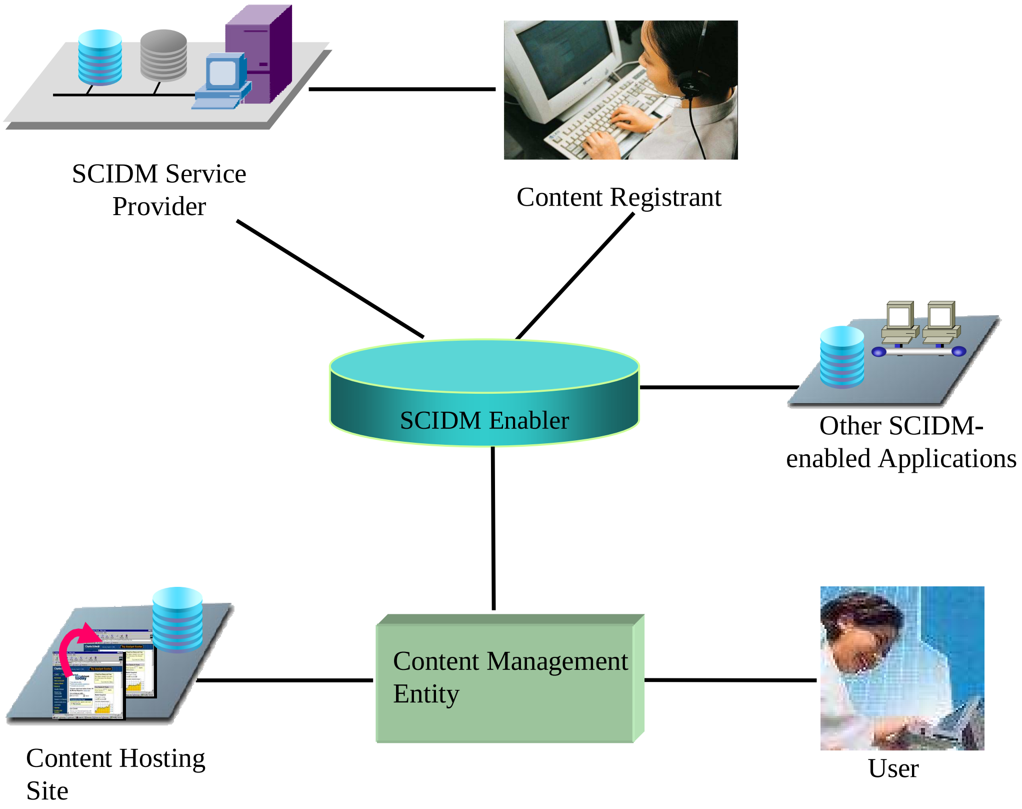

Figure 2: SCIDM enabler ecosystem

The main actors and their roles are described in the following.

- **SCIDM Service Provider**: Entity that operates SCIDM-enabled content identification and management service. Can also act as content registrant in some applications.

- **Content Registrant**: those who submit content to SCIDM enabler for content registration. May be different parties according to the applications, e.g, Content Provider in copyright protection applications, operator administrator or ordinary user in content filtering.

- **SCIDM-enabled Application**: an application that uses SCIDM enabler for content identification.

- **Content Hosting Site:** Entity willing to host a content item uploaded by a Content Provider or a User in compliance with some management policy, e.g. copyright protection policy. The management policy may be implemented by the Content Management Entity hosted in the site.

- **User:** Entity that pulls or be pushed some content, upon which the Content Management Entity may perform some management operations.

- **Content Management Entity (CME)**: An entity that bears some kind of content management function, e.g., copyright protection, infringement content filtering, monetization, and other content management, and communicates with the CIM via SCIDM Client for content identity query, content verification, and content registration. CME is part of a SCIDM-enabled Application and can be hosted in content hosting site, user device, network gateway, or other likely entity according to different applications.

5. SCIDM Enabler Description (Informative)

The SCIDM enabler specifies a set of secure content identification mechanisms for interoperable, persistent content identification and authentication. Secure identification and authentication of digital content would allow secure content transactions between all relevant entities (e.g., Content Provider, content distributor, service provider, operator, enabler, end user) in the service environment, resulting in a more trustworthy and efficient service/transaction environment. This will greatly benefit all parties involved, and enable various types of potential new services.

The SCIDM enabler provides the following main functionalities.

Content Registration: the enabler supports the registration of content and its Content Metadata by a trusted entity (e.g., a Content Provider). The enabler will assign a unique Content ID upon registration, and may issue a Content ID Certificate. Once the content is registered, it can be securely identified subsequently.

Secure Content Identification: the enabler supports the secure and persistent identification of various contents that have been registered. A modified content (which may or may not have been registered and assigned a new ID), may also be securely identified if their original has been registered. Upon a successful identification, the enabler returns the Content ID as well as some associated Content Metadata of the content, which can then be used to facilitate subsequent content management and processing.

Content Query: the enabler supports flexible ways to query for the content and its Content Metadata.

Security: the enabler is provisioned with security mechanisms for entity authentication and authorization, data integrity protection, and/or confidentiality protection.

The enabler can complement some other OMA enablers, for example, to be used to avoid the categorizing process being repeatedly applied to the same content submitted by different entities in CBCS [OMA-CBCS-RD], or to provide context for contextual advertising in Mobile Ad [OMA-MobAd-RD].

6. Requirements (Normative)

6.1 Modularisation

Content Registration: this functional module supports the registration of content and its Content Metadata. Once the content is registered, it can be securely identified. This is a mandatory functional module.

Content Identification: this functional module supports the identification of various contents that have been registered. A repurposed/modified content, if their original has been registered, may also be securely identified. This module may return the Content ID as well as some associated Content Metadata, which can then be used to facilitate subsequent content management and processing. This is a mandatory functional module.

Content Query: this functional module provides flexible ways to query for the content item and its Content Metadata. The query may be the Content ID, Content Fingerprint, or other Content Metadata. The query may return a Content ID, attributes associated with the queried content item, and/or any result of a computation that was triggered by the query. Through query, information on the content item can be acquired if the content item has been previously registered. This is a mandatory functional module.

Some requirements are intended to affect all the functional modules, and therefore are marked in the functional module column of the requirement's table as “General”.

6.2 High-Level Functional Requirements

6.2.1 Registration

Label	Description	Enabler Release	Functional module
SCIDM-REG-001	The SCIDM enabler SHALL be able to register an unknown content.	SCIDM 1.0	Content Registration
SCIDM-REG-002	The SCIDM enabler SHALL be capable upon registration to associate a unique Content ID and metadata to a content item.	SCIDM 1.0	Content Registration
SCIDM-REG-003	The SCIDM enabler SHALL be capable of issuing a Content ID Certificate and associating it with a content item.	SCIDM 1.0	Content Registration
SCIDM-REG-004	The SCIDM enabler SHALL be capable of associating a globally unique Content ID with the content to be used by other enablers/systems.	SCIDM 1.0	Content Registration
SCIDM-REG-005	The SCIDM enabler SHALL be capable of associating a local (i.e. to be used in different domains) Content ID which complements the global unique ID.	SCIDM 1.0	Content Registration
SCIDM-REG-006	The SCIDM enabler SHALL be capable of verifying the credential for content registration provided by a proxy entity (e.g. acting on behalf of a trustable website).	SCIDM 1.0	Content Registration

SCIDM-REG-007	The SCIDM enabler SHALL be able to support and differentiate between the registration request from different Principals (e.g. individual user, from proxy agent, or from traditional Content Provider).	SCIDM 1.0	Content Registration
SCIDM-REG-008	The SCIDM enabler SHALL support the content and associated metadata registration for different applications.	SCIDM 1.0	Content Registration
SCIDM-REG-009	The SCIDM enabler SHALL be capable of allowing the SCIDM Client to know how the authentication and verification of the Content Metadata is done during registration.	SCIDM 1.0	Content Registration
SCIDM-REG-010	The SCIDM enabler SHALL reuse the Content ID assigned by existing ID systems.	SCIDM 1.0	Content Registration
SCIDM-REG-011	The SCIDM enabler SHALL use Content Fingerprint for content registration.	SCIDM 1.0	Content Registration
SCIDM-REG-012	The SCIDM enabler SHALL support attributes to address the setting of content management rules (e.g., copyright control rules) on the content during content registration.	SCIDM 1.0	Content Registration
SCIDM-REG-013	The SCIDM enabler SHALL be capable of supporting the judgement of the trustworthiness of the registered metadata. Note: this is necessary as some of the registered metadata is later retrieved and used by other applications. For example, one way to address this is for SCIDM enabler to provide information about the source that registers the content and the metadata.	SCIDM 1.0	Content Registration
SCIDM-REG-014	The SCIDM enabler SHALL allow the registration of repurposed content.	SCIDM 1.0	Content Registration
SCIDM-REG-015	The SCIDM enabler SHALL support representing the relationship between the re-purposed content and the original registered content when registering the repurposed content.	SCIDM 1.0	Content Registration

Table 1: High-Level Functional Requirements – Registration items

6.2.2 Identification

Label	Description	Enabler Release	Functional module
SCIDM-IDEN-001	The SCIDM enabler SHALL be capable of securely/uniquely identifying registered content (including user generated content).	SCIDM 1.0	Content Identification
SCIDM-IDEN-002	It SHALL be possible for the SCIDM enabler to determine if a Content ID Certificate is associated with a content item.	SCIDM 1.0	Content Identification
SCIDM-IDEN-003	The SCIDM enabler SHALL support validating the Content ID Certificate.	SCIDM 1.0	Content Identification
SCIDM-IDEN-004	The SCIDM enabler SHALL support identifying content by content ID value.	SCIDM 1.0	Content Identification
SCIDM-IDEN-005	The SCIDM enabler SHALL support identifying content by Content Metadata (e.g., MD5 value of the content.)	SCIDM 1.0	Content Identification

SCIDM-IDEN-006	The SCIDM enabler SHALL support using Content Fingerprint for identification.	SCIDM 1.0	Content Identification
SCIDM-IDEN-007	The SCIDM enabler SHALL support querying the CIM to identify a content item and its attributes.	SCIDM 1.0	Content Identification and Content Query
SCIDM-IDEN-008	The SCIDM enabler SHALL support, upon query, verifying that the content (e.g., application software) matches its claimed attributes.	SCIDM 1.0	Content Identification
SCIDM-IDEN-009	The SCIDM enabler SHALL support identifying content by leveraging identification history.	SCIDM 1.0	Content Identification
SCIDM-IDEN-010	The SCIDM enabler SHOULD allow the CIM to support identifying content by Digital Watermark. Note: “SHOULD” is used to avoid enforcing the implementation of digital watermarking techniques in CIM, although it may be desirable in some applications.	SCIDM 1.0	Content Identification
SCIDM-IDEN- 011	The SCIDM enabler MAY allow the SCIDM Client to support identifying content by Digital Watermark. Note: “MAY” is used to avoid enforcing the implementation of digital watermarking techniques in SCIDM Client.	SCIDM 1.0	Content Identification
SCIDM-IDEN-012	The SCIDM enabler SHALL be capable of securely identifying all registered single media objects.	SCIDM 1.0	Content Identification
SCIDM-IDEN-013	The SCIDM enabler SHOULD be capable of securely identifying compound objects which are registered as a whole or have their components registered. Note: This requirement is optional as identifying compound objects is expected to be more complex as it has to deal with the relationship of individual objects within a compound object.	SCIDM 1.0	Content Identification
SCIDM-IDEN-014	The SCIDM enabler SHALL have a flexible framework to allow using different types of fingerprint algorithms for different types of contents	SCIDM 1.0	Content Identification
SCIDM-IDEN-015	The SCIDM enabler SHALL have a flexible framework to support, by providing tailored metadata, content identification for different applications, e.g. copyright verification, pornographic content filtering, etc.	SCIDM 1.0	Content Identification
SCIDM-IDEN- 016	The SCIDM enabler SHALL support leveraging different identification mechanisms that meet the security requirement and achieve the best performance.	SCIDM 1.0	Content Identification
SCIDM-IDEN- 017	The SCIDM enabler SHALL allow the SCIDM Client to select the identification mechanisms.	SCIDM 1.0	Content Identification
SCIDM-IDEN- 018	The SCIDM enabler SHALL support allowing the SCIDM Client to know the security implication of each identification mechanism.	SCIDM 1.0	Content Identification
SCIDM-IDEN-019	The SCIDM enabler SHALL allow SCIDM Clients to report errors (e.g. attribute inconsistency) during content identification and query	SCIDM 1.0	Content Identification
SCIDM-IDEN-020	The SCIDM enabler SHALL allow the SCIDM Client to set priority in terms of processing time in its content query and identification request.	SCIDM 1.0	Content Identification

SCIDM-IDEN-021	The SCIDM enabler SHALL allow the SCIDM Client to provide information about the application (which is performing the identification) to CIM for priority disposition of the requests.	SCIDM 1.0	Content Identification
SCIDM-IDEN-022	The SCIDM enabler SHOULD be capable of identifying the repurposed content as a derivative of the original registered content. Note: "SHOULD" is used instead of "SHALL" as the capability depends on to what degree the content has been repurposed.	SCIDM 1.0	Content Identification
SCIDM-IDEN-023	The SCIDM enabler SHALL support the representation of the relationship between the re-purposed content and the original registered content when identifying the repurposed content	SCIDM 1.0	Content Identification
SCIDM-IDEN-024	The SCIDM enabler SHOULD provide support for identification of content based on partial content.	SCIDM 1.0	Content Identification

Table 2: High-Level Functional Requirements – Identification items

6.2.3 Query

Label	Description	Enabler Release	Functional module
SCIDM-QUER-001	The SCIDM enabler SHALL provide a flexible framework to allow querying the registered content database by some of the content's attributes (e.g., content ID, Name, or Content Fingerprint).	SCIDM 1.0	Content Query
SCIDM-QUER-002	The SCIDM enabler SHALL be capable of providing content management rules as attribute to SCIDM Client.	SCIDM 1.0	Content Query
SCIDM-QUER-003	The SCIDM enabler SHALL enable the following responses to a query (non exhaustive list): - content IDs - Content Metadata associated with queried content items - results of computations that are triggered by the queries (e.g. policy-based decisions).	SCIDM 1.0	Content Query

Table 3: High-Level Functional Requirements – Query items

6.2.4 Security

Label	Description	Enabler Release	Functional module
SCIDM-SEC-001	The CIM SHALL detect replay attacks	SCIDM 1.0	General

SCIDM-SEC-002	The SCIDM enabler SHOULD take necessary measures to help prevent the denial of service attacks. Informational note: SHOULD is used instead of SHALL as this is not the core functionality of the enabler, and it can be achieved by other implementation specific means.	SCIDM 1.0	General
SCIDM-SEC-003	The SCIDM enabler SHALL be able to securely identify which CIM provided a particular Content ID.	SCIDM 1.0	General
SCIDM-SEC-004	The SCIDM enabler SHALL be able to associate information to a given Content ID in order to characterize it (e.g. public, private, permanent, temporary, etc)	SCIDM 1.0	General

Table 4: High-Level Functional Requirements – Security Items

6.2.4.1 Authentication

Label	Description	Enabler Release	Functional module
SCIDM-AUTH-001	The SCIDM enabler SHALL allow the CIM to authenticate SCIDM Clients before content registration.	SCIDM 1.0	Content Registration
SCIDM-AUTH-002	The SCIDM enabler SHALL allow all SCIDM Clients to authenticate the CIM.	SCIDM 1.0	General
SCIDM-AUTH-003	The SCIDM enabler SHALL support the mutual authentication between CIMs when synchronizing registration information.	SCIDM 1.0	Content Registration
SCIDM-AUTH-004	The SCIDM enabler SHALL be able to perform anonymous queries	SCIDM 1.0	Content Query

Table 5: High-Level Functional Requirements – Authentication Items

6.2.4.2 Authorization

Label	Description	Enabler Release	Functional module
SCIDM-AUTR-001	The SCIDM enabler SHALL be able to authorize access to the CIM functionalities to authenticated SCIDM Clients	SCIDM 1.0	General

Table 6: High-Level Functional Requirements – Authorization Items

6.2.4.3 Data Integrity

Label	Description	Enabler Release	Functional module
SCIDM-INTE-001	The SCIDM enabler SHALL support data integrity protection between the CIM and the SCIDM Clients.	SCIDM 1.0	General
SCIDM-INTE-002	The SCIDM enabler SHALL support data integrity protection between CIMs when synchronizing the registration information.	SCIDM 1.0	Content Registration

Table 7: High-Level Functional Requirements – Data Integrity Items

6.2.4.4 Confidentiality

Label	Description	Enabler Release	Functional module
SCIDM-CONF-001	The SCIDM enabler SHALL allow CIMs to perform a confidentiality protected communication with the SCIDM Client.	SCIDM 1.0	General
SCIDM-CONF-002	The SCIDM enabler SHOULD allow SCIDM Clients to perform a confidentiality protected communication with the CIM. Informative note: SHOULD is used instead of SHALL as some applications will not need such level of confidentiality	SCIDM 1.0	General
SCIDM-CONF-003	The SCIDM enabler SHALL allow CIMs to perform a confidentiality protected communication with other CIMs	SCIDM 1.0	General

Table 8: High-Level Functional Requirements – Confidentiality Items

6.2.5 Charging

Label	Description	Enabler Release	Functional module

Table 9: High-Level Functional Requirements – Charging Items

6.2.6 Administration and Configuration

Label	Description	Enabler Release	Functional module
SCIDM-ADM-001	The SCIDM enabler SHALL support synchronization of information among CIMs (e.g. content registration information).	SCIDM 1.0	General

Table 10: High-Level Functional Requirements – Administration and Configuration Items

6.2.7 Usability

Label	Description	Enabler Release	Functional module
SCIDM-USAB-001	The delay experienced by the User during content uploading and downloading SHALL be comparable to the case in which SCIDM is not used.	SCIDM 1.0	Content Identification and Content Query

SCIDM-USAB-002	The SCIDM enabler SHALL enable the SCIDM Client to carry out content identification also when the SCIDM Client cannot utilize network connectivity.	SCIDM 1.0	Content Identification
SCIDM-USAB-003	The SCIDM enabler SHALL enable the SCIDM Client to carry out content query also when the SCIDM Client cannot utilize network connectivity.	SCIDM 1.0	Content Query
SCIDM-USAB-004	The SCIDM enabler SHALL NOT block the content to be consumed in devices which do not implement the SCIDM enabler.	SCIDM 1.0	General

Table 11: High-Level Functional Requirements – Usability Items

6.2.8 Interoperability

Label	Description	Enabler Release	Functional module

Table 12: High-Level Functional Requirements – Interoperability Items

6.2.9 Privacy

Label	Description	Enabler Release	Functional module

Table 13: High-Level Functional Requirements – Privacy Items

6.3 Overall System Requirements

Label	Description	Enabler Release	Functional module
SCIDM-SYST-001	It SHALL be possible to support distributed CIM mechanism to address system scalability.	SCIDM 1.0	General

Table 14: High-Level System Requirements

Appendix A. Change History (Informative)

A.1 Approved Version History

Reference	Date	Description
n/a	n/a	No prior version –or- No previous version within OMA

A.2 Draft/Candidate Version 1.0 History

Document Identifier	Date	Sections	Description
Draft Versions OMA-RD-SCDIM-V1_0	07 Mar 2008	all	Skeleton created
	07 Apr 2008	1	Scope text added
	25 Apr 2008	all	Introduction, first requirements, use cases, and definitions added. While also changing to new RD Template – e.g. new Use Cases now to Appendix.
	22 Jun 2008	4, 6, B2	Content Life Cycle description (4), new use case (B2), 2 new general reqs
	10 Jul 2008	3, 6, B	3 use cases added, with corresponding requirements
	10 Aug 2008	3, 4, 6, B	Additional introduction text, use cases update, requirements, 1 definition,
	29 Aug 2008	2, 3, 4, 5, 6	Requirements, references, definitions added. Enabler description updated.
	13 Oct 2008	all	Editorial cleanups
	23 Oct 2008	all	All RDRR resolutions except 2 agreed CRs
	27 Oct 2008	4,5,6	Separating requirement (tables) wrt modules, changing attribute -> metadata
	28 Oct 2008	all	Minor editorial cleanups
Candidate Version OMA-RD-SCIDM-V1_0	16 Dec 2008	n/a	Status changed to Candidate by TP TP ref# OMA-TP-2008-0445R01-INP_SCIDM_V1_0_RD_for_Candidate_approval

Appendix B. Use Cases (Informative)

B.1 Content Registration

B.1.1 Short Description

In order for the content to be identifiable by the SCIDM enabler, it should be first registered to the CIM. Different entities including traditional Content Provider, CIM's administrator, ordinary user, etc, may provide content for registration. Upon the registration, the CIM should try its best to assure the trustworthiness of the registered metadata associated with the content. Different applications may require different levels of trustworthiness of the metadata.

For example, for the copyright protection application, the copyright ownership is the most important metadata. If the content is owned by traditional Content Provider, formal proof of ownership should be presented. Then the CIM can verify the proof by manual manner or save the proof as metadata for users of SCIDM enabler to judge by themselves. If the content is owned by ordinary user, it is somewhat difficult to assure the trustworthiness of the claimed ownership. User generated content (UGC) registration will be discussed in B.7.

For content filtering and other applications, some automatic or manual methods may be used for content registration. For example, upon the registration of a SMS text or image for filtering, key word or character matching can be done to verify if the content submitted for registration really needs to be filtered. For most cases, maybe only manual approaches for metadata verification are feasible / available during registration. These manual approaches are out of scope of SCIDM.

As a general flow, the registration will proceed as follows. Some party provides some content items to CIM for registration. CIM first authenticates the party and then verifies the metadata by available and acceptable means. Then CIM allocates content ID, and extracts the Content Fingerprint. Optionally, CIM can create the Content ID Certificate containing the Content ID and the necessary metadata for the proposed application. For users of SCIDM enabler to judge the trustworthiness of the metadata by themselves, CIM records information about the party who registers the content, metadata verification methods, etc. Finally, CIM notifies the registration party the registration result.

B.1.2 Market Benefits

Content registered to CIM can be managed in a trustworthy way for various applications.

B.2 Leveraging Multiple Identification Mechanisms

B.2.1 Short Description

There are several means to identify content, e.g., by Content ID value, by content metadata, by Digital Watermark, by Content Fingerprint, etc. Some of these methods are not necessarily secure and can be circumvented with different levels of difficulty, while Content Fingerprint-based approach is most difficult to compromise and is most secure. Yet considering the actual service environment, the methods not so secure may be acceptable in some applications and scenarios. For example, to identify the copyright-infringement content in the content-sharing websites, or to identify the content for advertisement association, if finding the matching content in CIM by the Content ID value□metadata or the extracted watermark from the content to be identified is successful, the result is to some extent creditable and may be accepted. If finding the matching content by these “non-secure” means is not successful, then the Content Fingerprint-based method can be tried. If it also fails, it can then be concluded that the content has not been registered with the CIM.

Another situation in actual service environment is that some Content ID value and metadata can be changed purposely by someone, so that identification by these means will fail. For these contents, if the Content Fingerprint-based approach can identify the content successfully, then the modified ID, metadata and this identification record (referred to as the identification history) can be used later in the identification of the same content. So Content Fingerprint may not always need to be used when identifying the same content.

Here is an example flow. When the Content Management Entity needs to verify a piece of content, it selects the appropriate identification means based on its performance requirement and security requirement, and then requests the CIM to identify the content using the selected means. If successful, the CIM returns the result to the Content Management Entity and the Content Management Entity can perform its managements on the content. If it fails and Content Fingerprint has not been used, the Content Management Entity can choose to request the CIM to identify the content based on the Content Fingerprint.

B.2.2 Market Benefits

Content Management Entity residing in website, service gateway or user device, and CIM can identify content with higher performance (less resource and time), therefore reducing the investment for content identification.

B.3 Copyright Verification in Posting

B.3.1 Short Description

Posting of user generated content has become increasingly popular. In this use case, content hosting site identifies the content a user would like to upload and makes sure it has the right to post the content.

A User requests to upload a content item to a Content Hosting Site. Before allowing the posting of the content, the Content Hosting Site generates the Content Fingerprint using the Content Fingerprint Extractor, and verifies the Content ID Certificate if available. If verification is successful, then the ID and potentially the copyright information included in the Content ID certification can be used by the Content Hosting Site as discussed below. If verification fails or no certificate is available, the Content Hosting Site queries the CIM by sending the generated Content Fingerprint to identify the content in the CIM’s database. If the content is found in the database, the CIM returns the content ID; otherwise, the CIM can choose to register the content in its database, assign a content ID, and issue a Content ID Certificate. The Content Hosting Site uses the Content ID to check if the content is in a database of copyrighted content. If yes, it may choose to acquire a license from the Content Provider to post the content, or refuse to post the content; if not copyrighted, then the hosting site may choose to post the content.

In the above process, if the content is found in the database and the copyright information is available in CIM, then the CIM may return the Content ID together with the copyright information for the use of the Content Hosting Site.

B.3.2 Market Benefits

User can reduce the risk of getting into legal complication as a result of uploading and posting a copyrighted content. Content Hosting Site can avoid any legal hassles by making sure it is not violating the copyright law by illegally hosting a copyrighted content. CIM generates revenue by providing the secure content identification service. Content Fingerprint Extractor generates revenue by providing a valuable service to Content Hosting Site, CIM or other entities. Content Provider (CP) can get better protection of its valuable content and minimize revenue loss.

B.4 Content Identification for Filtering

B.4.1 Short Description

Contents that may infringe the fair use of the Internet and mobile services are frequently observed, e.g., SMS/MMS spam, and pornographic. To correctly identify these contents and then filter or block them according to some rules is very important. Those contents that need to be filtered should first be registered with the CIM, and Content Fingerprints are extracted and recorded together with some other attributes. The entity to register the content with the CIM can be user application, service provider, operator, etc. When the entity that is responsible for content monitoring and filtering receives a content item, it will contact the CIM for content identification and may retrieve attributes from CIM to help the filtering process. The entity for content monitoring and filtering can be, for example, service gateway, server, or user device.

A flow is described here for illustration purpose. A User who is a child requests to download a piece of content from a website. After the content is downloaded into the user's device, the Content Management Entity in the device checks if the content needs to be controlled according to the rules set by the child's parents. So it generates the Content Fingerprint, and verifies the Content ID Certificate if it is available. If verification fails or no certificate is available, the Content Management Entity queries the CIM by sending the generated Content Fingerprint or the content itself to identify the content in the CIM's database. If the content is found in the database, the CIM returns the content ID. The Content Management Entity uses the Content ID to check the category of the content in a content category database. If it exists, it may retrieve the category information and judge if it should block the content.

B.4.2 Market Benefits

User will receive fine content and is shielded to a large extent from infringement content, and has the option to complain about the infringement content when receiving it for later filtering. Content Management Entity can correctly identify the infringement content and filter it, so as to satisfy the requirements from user, operator, service provider, etc.

CIM generates revenue by providing the secure content identification service.

B.5 Content Verification for Anti- Virus and Malicious Plug-in

B.5.1 Short Description

Nowadays many virus and malicious plug-in’s are wrapped in some types of content, e.g. software, and intrude into user’s device when the software is downloaded and installed. The anti-virus software has limited effect since it is easy to have many new variations of the virus. Downloaded software thus needs to be verified before it is installed.

Alice downloads a software from a software portal. After the download is completed, some Content Management Entity that may reside in Alice’s device or the gateway automatically calculates the fingerprint of the software and sends the fingerprint and other necessary information, e.g. software name, version, ID, to the Content Identity Manager (CIM). CIM checks if the software is registered in the database and if the associated metadata is consistent with the registered one and returns the result to Alice’s device. If the result is not a match, the software is not allowed to be installed or executed, and some prompt is presented to Alice.

In the above example, the Content Management Entity can also query the CIM by sending the name, ID, and version of the software. After the content is found in the database, the CIM returns the fingerprint in the database to the Content Management Entity. Content Management Entity checks if the returned fingerprint matches the one it generated, and takes appropriate actions according to the checking result.

B.5.2 Market Benefits

User gets clean content, and is shielded from virus and malicious plug-in’s. Content Management Entity can correctly verify if the downloaded content is clean, and avoid the user device being infected by virus and malicious plug-in’s. It generates revenue by providing a valuable service to user, software vendor, etc. CIM generates revenue by providing the secure content identification service

B.6 Copyright Control in P2P Networks

B.6.1 Short Description

P2P service has become very popular in the Internet and mobile environment. Unfortunately a lot of copyright infringement contents are transmitted and shared in P2P networks. To correctly identify the copyrighted content being transmitted is a desire for copyright owners.

A use case is described here. Contents are being transmitted in the P2P environment; some are copyright infringement and others not. There is one Content Management Entity residing in the gateway where the P2P flow will have to pass through. The Content Management Entity will make a copy of every piece of content passing through it and verify, with the help of the SCIDM enabler, if the content is copyrighted. The Content Management Entity may choose to only record some pieces/parts of the content. It should send sufficient number of pieces to CIM for the identification of the content and copyright control. According to the results and the control rules, the Content Management Entity takes some actions, e.g. record the transmission time and parties involved in the transmission, or block the P2P flow.

B.6.2 Market Benefits

Content Management Entity can correctly identify the infringement content in the P2P networks and filter it, so as to satisfy the requirements from user, operator, service provider, etc.. CIM generates revenue by providing the secure content identification service. Content Provider: can get better protection of its valuable content and minimize revenue loss.

B.7 Registration of User Generated Content

B.7.1 Short Description

In content registration by traditional Content Provider, the content owner needs to provide a certificate of the content which is issued by an authority to show the copyright ownership. But it is generally difficult and complicated for an individual user to obtain the copyright certificate for contents created by the user such as personal photo, video, etc. So, for an individual user who would like to get protection of the potential benefit of his/her works, namely User Generated Content (UGC), the traditional registration method is not suitable.

In some service environment, the service may be familiar with his users, such as those who have been using the service for a long time. So the service can provide some assurance of the user's ownership to his content. For example, a user has used an original picture sharing website for a long time, and gets to be well-known and trusted by the site. Then the website may be willing to act as a proxy for its users to register their content to the CIM in exchange for service charge or revenue sharing with the user.

If the user wishes to register his content to CIM using the service as his proxy, he may first have some agreement with the service. Then when the user wants to register his content, he needs to obtain some credential information created by the service to assert the user's ownership to the content. After generating successfully the credential information for the UGC, the credential material could be fed back to the user, so that the user could register his/her UGC-content together with the credential to the CIM. Alternatively, upon request by the user, the proxy agent of the website could deliver directly the UGC-content and the ensuring information to the CIM for registration. If the CIM trusts the service, it verifies the credential information and gets some assurance of the ownership information. Whether successful or not, the CIM would respond with the result of registration to the user (directly or indirectly).

B.7.2 Market Benefits

The user and website could get revenue from the registered user generated content, because of the potential value of the UGC. The UGC creation will be much more flourish and the quality of UGC will be improved.

For the CIM, its identification service can be used more widely and generate more revenue.

Appendix C.

N/A.