Secure Content Identification Mechanism Architecture

Candidate Version 1.0 – 28 Jul 2009

Open Mobile Alliance

OMA-AD-SCIDM-V1_0-20090728-C

Contents

Figures

1. Scope (Informative)

This document describes the architecture of the SCIDM Enabler. The architecture is based on the requirements and the use cases included in SCIDM Requirements document [SCIDM-RD] and described at high level as believed to be significant from the architectural point of view.

The objective of the SCIDM Enabler is to leverage available identification mechanisms to identify all kinds of content including both premier and user generated content for various applications. Among the available identification mechanisms, Content Fingerprint based identification is mandatory to be supported in this Enabler. The architecture shown in this document is intended to facilitate the development of specifications for content registration, query and identification. The description of the architecture comprises the definition of functional components and the interfaces used or exposed by these functional components.

Note that this Enabler does not specify application-specified mechanisms, such as the content management schemes and rules. Definition of metadata for specific applications is also out of scope.

2. References

2.1 Normative References

[OSE] "OMA Service Environment", Open Mobile Alliance™, URL:http://www.openmobilealliance.org/

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

[SCIDM-RD] "Secure Content Identification Mechanism Requirements", Open Mobile Alliance™, OMA-RD-SCIDM-V1_0, URL:http://www.openmobilealliance.org/

[ISO-MPEG-7] *ISO/MPEG N4674, Overview of the MPEG-7 Standard*, v 6.0, J.M. Martínez, ed., MPEG Requirements Group, Jeju, Mar. 2002

[MPEG-7-IMG-SIG] "ISO/IEC 15938-3:2002/Amd.3 Image Signature Tools," Mar. 2009

[RFC2141] R. Moats, "URN Syntax", RFC 2141, May 1997.

[OMA-SEC_CF-AD] "Application Layer Security Common Functions Architecture", Open Mobile Alliance™, OMA-AD-SEC_CF-V1_0, URL:http://www.openmobilealliance.org/

[OMADICT] "Dictionary for OMA Specifications", Version 2.7, Open Mobile Alliance™, OMA-ORG-Dictionary-V2_7, URL:http://www.openmobilealliance.org/

2.2 Informative References

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.

3.2 Definitions

Compound Content	A content item that consists of multiple individually recognizable content objects.
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 a content item that can uniquely identify the content item.
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 of a content item.
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.
Registration Client	A SCIDM Client that makes requests to the CIM for content registration
Identification Client	A SCIDM Client that makes requests to the CIM for content identity query and content verification

3.3 Abbreviations

AD	Architecture Document
CIM	Content Identity Manager
CntIDCert	Content ID Certificate
ID	IDentity
OMA	Open Mobile Alliance
SCIDM	Secure Content IDentification Mechanism

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 more convenient, 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.

The SCIDM Enabler provides a basic service: a standardized trusted content ID (which is a more efficient, and secure (trusted) representation of the content than simply a content name). Specific applications/services can take advantage of SCIDM to offer a more efficient and trustworthy service.

The objective of this document is to describe the architecture for Secure Content Identification Mechanism Enabler.

4.1 Version 1.0

It is planned to meet all the SCIDM requirements [SCIDM-RD] in this release. No future releases are currently planned.

5. Architectural Model

The SCIDM Enabler conforms to the OMA Service Environment [OSE].

5.1 Dependencies

The SCIDM enabler depends on the following enabler or specifications

- OMA SEC_CF Enabler [OMA-SEC_CF]
- ISO/IEC MPEG-7 [ISO-MPEG-7]
- ISO/IEC MPEG-7 Image Signature Tools [MPEG-7-IMG-SIG]
- Uniform Resource Names (URN) [RFC2141]

5.2 Architectural Diagram

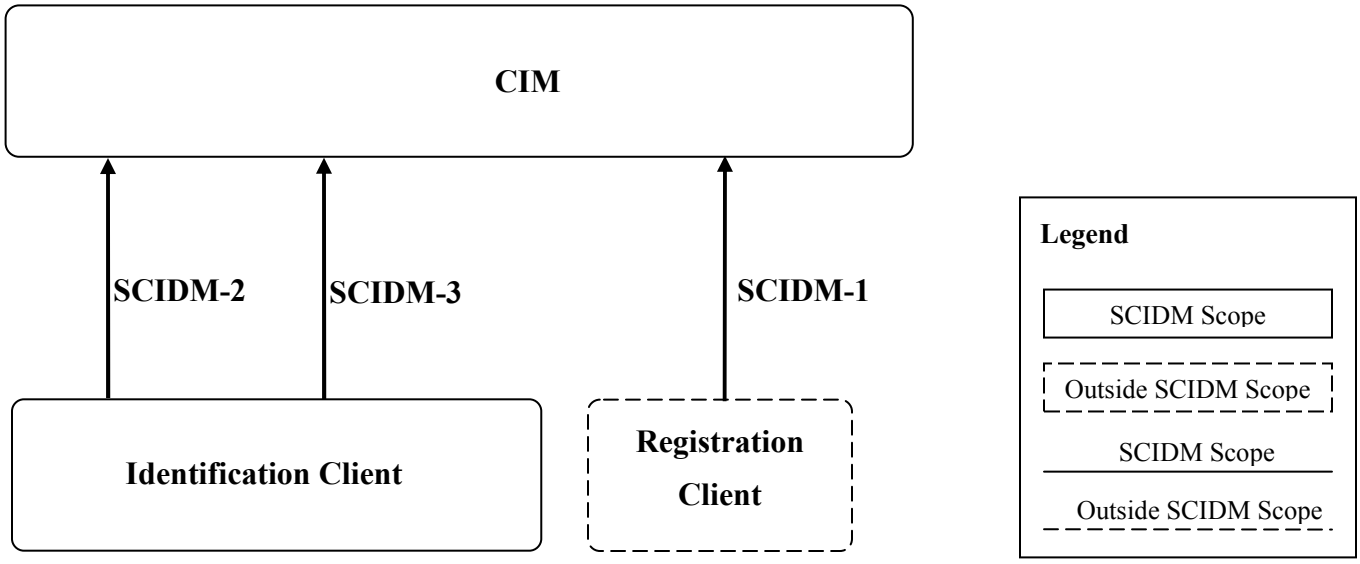

Figure 1: SCIDM architectural diagram

5.3 Functional Components and Interfaces/reference points definition

5.3.1 SCIDM Enabler Functional Components

5.3.1.1 CIM

CIM acts as the SCIDM Server residing in the network, in charge of managing content registration, responding to content identity query verification requests, and issuing Content ID Certificates.

The CIM performs the following functions: .

- Content Registration Function
- Identification and Query Function
- Fingerprint Handle Function

5.3.1.1.1 Content Registration Function

The Content Registration Function performs the following actions:

- Receive content registration requests from authorized principals such as the Registration Client.
- Judge the trustworthiness of the content metadata submitted. Trustworthiness of the content metadata is key for content identification and corresponding management actions. The CIM SHOULD leverage reasonable mechanisms to judge the trustworthiness. Some mechanisms MAY be done manually or dependent on the specific application, which is out of scope of this Enabler. The following common actions SHALL be supported by this enabler:
    - Record the authentication mechanism used to authenticate the Registration Client when the content is registered and upon request, make it available to the Identification Client.
    - Differentiation of different principals, such as traditional content provider, individual user, proxy agent, the operator of the Enabler, etc. The differentiation is based on the validated principal type provided by the Content Registrant. When the Content Registrant registers itself to CIM for subscribing to the service of SCIDM, the CIM SHALL requests the Content Registrant to provide its principal type. Depending on the deployment, the CIM SHOULD try to validate the provided principal type through available mechanisms. It is recommended that the CIM SHOULD validate the principal by out-of-band mechanisms if the provided principal type is traditional content provider, or proxy agent, or operator of the Enabler. If the validation fails, the registration for subscribing to the SCIDM service will be refused. If the provided principal type is individual user, some identity information MAY be provided and validation by out-of-band mechanisms MAY be done. Without these identity information and validation, the registration for subscribing to the SCIDM service from individual user MAY also be accepted, but with a low level of trustworthiness. These steps in service subscription are out of the scope of the SCIDM Enabler.
    - Handling the credentials of the registered metadata provided by a proxy entity (e.g. acting on behalf of a trustable website) SHALL be supported. If the Registrant is an individual user, e.g., for User Generated Content, in most cases this scheme MAY be used. This does not mean all registration from an individual user MUST use this scheme, however the metadata registered using this scheme MAY receive higher level trust than those not using this scheme.
    - Record the level of trustworthiness if different levels of trust are used.
    - Record the mechanism(s) used to verify the trustworthiness and upon request, make it available to the Identification Client.
- Extract the content fingerprint with the help of the Fingerprint Handle Function as defined in Section 5.3.1.1.3. If the content is compound, split it into single objects, and extract the fingerprint of each object.
- Examine if the content has been registered before based on the extracted fingerprint, with the help of Fingerprint Handle Function. Repurposed content SHOULD be specially treated. Some repurposed content item MAY be

treated as the same as the original content, some MAY NOT, which depends on the application and/or policy deployed in the CIM which is out of scope of this Enabler.. If the same content item has been registered, decline the registration.

- Allocate globally unique ID for the Content. If the registered content item is a compound one, and there are no objects in this compound content item is found already registered, in addition to the ID for the compound one, allocate an ID for each object in it. If there are objects in the compound content item is found already registered, in addition to allocate an ID for each unregistered object, whether or not allocate an ID for the compound content item is determined by the policy deployed in the CIM.
- Optionally create Content ID Certificate for the Content
- Store the content ID, Certificate and metadata. If the content item is compound, the IDs for individual objects are linked to the ID for the compound one.
- Store the Content Fingerprint and its ID with the help of the Fingerprint Handle Function. If the content item is compound, store this information for each object.

5.3.1.1.2 Identification and Query Function

The Identification and Query Function is responsible for interacting with the Identification Client for content identification and query. It SHALL support online identification and provide service to facilitate offline identification in the client side. The online identification SHALL include the following actions:

- Deal with the content identification and/or query request from the Identification Client
- Select the appropriate content identification mechanisms based on the application type, content type and client's preference. If the content item is compound, it SHALL be split into single objects, and then the identification mechanism for each object is selected based on its corresponding content type.
- Perform the identification by content ID if so selected.
- Perform the identification by content metadata if so selected.
- Perform the identification by content fingerprint with the help of the Fingerprint Handle Function if so selected. In this case, it SHALL be able to identify repurposed content and partial content if it is treated as the same as the original content and its corresponding original content is registered. Digital watermark is treated as one type of fingerprint.
- If matched content is found, retrieve the necessary metadata from the Content Metadata Database.

The Identification and Query Function also provides service to support offline identification at the client side. This SHALL include the following actions:

- Send notification to the Identification Client for content registration information synchronization. This action MAY be triggered by some policy deployed in the CIM. See Section 5.3.3.1.2 for more information about the policy for content registration information synchronization.
- Send content registration information, usually in the form of Content ID Certificate, to Identification Client upon its request for offline content identification.

5.3.1.1.3 Fingerprint Handle Function

The Fingerprint Handle Function is a special part in CIM. A CIM MAY have more than one Fingerprint Handle Functions, each for one type of content in one application scenario.

This Function performs the following actions:

- Extract the fingerprints of content upon the request from the Registration Function and Identification and Query Function.

- Store and retrieve the fingerprints from the fingerprint database, upon the request from the Identification and Query Function.

5.3.1.2 Identification Client

The Identification Client is used to interact with CIM for content identification.

The Identification Client performs the following functions:

- Identification Client Function
- CntIDCert Handle Function
- Fingerprint Extraction Function

5.3.1.2.1 Identification Client Function

The Identification Client Function performs both online content identification and offline content identification. For online content identification, the following actions are performed:

- Interact with CIM for content identification.
- Select preferred identification mechanisms.
- Collect necessary information of the content for content identification. If the selected identification mechanism is fingerprint, the Client Identification Function MAY send the content itself or the extracted fingerprint to the CIM. For efficiency or other considerations depending on the application, the Client Identification Function MAY send partial content or the fingerprint extracted from partial content as the identification information to the CIM. If the content is compound, the Client Identification Function SHALL collect relevant information for each member object.
- Report errors to CIM.

For offline content identification, the following actions are performed:

- Offline identify the content if so selected, with the help of the CntIDCert Handle Function. The offline identification is usually done with the Content ID Certificate which is managed by the CntIDCert Handle Function. So the Content ID Certificate needs to be obtained before the content can be identified offline. This MAY be achieved by the synchronization between the CIM and the client described as following, or delivering the Content ID Certificate along with the content, or other means that is not specified in the enabler.

  To perform offline identification, the Content ID Certificate SHALL be verified. The Content Fingerprint contained in the certificate is then used to check against the Content Fingerprint extracted from the content item. If matched, the metadata in the Content ID Certificate MAY be used to perform some management actions.
- Receive the notification sent by the CIM for content registration information synchronization.
- Synchronize the content registration information, usually in the form of Content ID Certificate, from the CIM for offline identification. This action MAY be triggered upon receiving the notification for content registration information synchronization from the CIM or by some policy deployed in the client. See Section 5.3.3.1.2 for more information about the policy for content registration information synchronization.

5.3.1.2.2 CntIDCert Handle Function

The CndIDCert Handle Function is responsible for the verification, storage, maintenance, retrieval of the Content ID Certificate. The format of the Content ID Certificate and the verification process SHALL be specified, and format extension SHALL also be supported. The storage, maintenance, and retrieval of the Content ID Certificate are out of the scope of the SCIDM Enabler.

5.3.1.2.3 Fingerprint Extraction Function

The Fingerprint Extraction Function is used to extract the fingerprint of content upon the request from the Identification Client Function. There MAY be more than one Fingerprint extraction functions residing in the Identification Client, each for one type of content in one application scenario. The SCIDM Enabler SHALL support various Content Fingerprinting algorithms, and some Content Fingerprinting algorithms will be specified. Note that one content item can potentially be registered multiple times using different content fingerprints extracted by different supported Content Fingerprinting algorithms, potentially by different registration clients. It is to CIM's advantage to bind them together. How to bind them together is out of the scope of this enabler though.

5.3.2 Other relevant functional components (Informative)

5.3.2.1 Registration Client

The Registration Client is used for content registration by the Content Registrant.

The Registration Client send content registration request to the CIM, and provide content metadata to the CIM, then receive the response from the CIM. It can send the content itself to CIM, and also can extract the fingerprint and send the fingerprint to CIM. The behavior of the Registration Client is out of the scope of the SCIDM Enabler.

5.3.3 Interfaces

5.3.3.1 SCIDM Enabler Interfaces

5.3.3.1.1 SCIDM-1

The SCIDM-1 interface is exposed by the CIM and can be used by any authorized principal, such as the Registration Client to submit content registration requests and get appropriate responses.

The parameters in the content registration request MAY include:

- Content itself or the content fingerprint, and in the latter case, the ID of the fingerprint extraction algorithm
- Content metadata, which depends on the application scenario.
- The information that proves the ownership of the Registration Client to the registered content.
    - E.g. if the Registration Client represents an individual user, a credential provided by a proxy entity (e.g. acting on behalf of a trustable website) SHOULD be provided, which can help the registered content metadata to receive an appropriate level of trust by the CIM.

The parameters in the content registration response SHALL include:

- The registration result code
- In some cases, the Content ID Certificate of the registered content.

5.3.3.1.2 SCIDM-2

The SCIDM-2 interface is exposed by the CIM and can be used by any authorized principal, such as the Identification Client, a remote CIM to submit content identification or content query requests, and get appropriate response.

The parameters in the content identification and query request MAY include:

- The content itself or the fingerprint extracted from the content, and in the latter case, the ID of the fingerprint extraction algorithm.
- Type of the application scenario in which the content is to be identified.

- Client's preference of content identification mechanisms, with some necessary information about the content.

The parameters in the content identification and query response SHALL include:

- The identification result code
- Requested content metedata for the specified application
- In some cases, Content ID Certificate.

5.3.3.1.3 SCIDM-3

The SCIDM-3 interface is exposed by the CIM. It can be used by any authorized principal, such as the Identification Client for content registration information synchronization. The authorized principal MAY use the information then for offline identification.The synchronization information MAY be delivered in the form of Content ID Certificate. The synchronization process can be triggered by some policy that MAY be pre-configured in the client or the CIM. The policy is usually configured in a way such that synchronization MAY help increase the efficiency and speed of the identification process. It has two kinds of flow. The first is that the Identification Client initiates to submit request to the CIM for synchronization. The second is that the CIM sends notification to the clients who have registered with the CIM previously, and then the client submits request to CIM.

In the following application scenarios, the synchronization process SHOULD be triggered to facilitate later offline identification:

- Spam filtering. Since spam (SMS/MMS spam, email spam, etc) is usually widespread, after the CIM gets new spam registered, it SHOULD notify the clients to synchronize them for offline identification.
- After a client finds that some content has been observed in high frequency, exceeding a predefined threshold for example, the client MAY initiate the synchronization process for subsequent offline identification.
- For some clients that are not often connected to the network, the clients MAY initiate the synchronization process when it is connected to the CIM.
- Other possible scenarios.

5.4 Security Considerations

5.4.1 Authentication and Authorization

Deployment of the SCIDM Enabler chooses whether to provide mutual authentication and authorization between the main SCIDM Enabler entities, such as between SCIDM Client and CIM. If it does so, only authorized principals will be able to perform SCIDM operations, such as content registration, content identification, content query and synchronizing registration information between CIMs, etc. A suitable authentication mechanism is specified in [OMA-SEC_CF-AD].

5.4.2 Integrity and Confidentiality

Deployment of the SCIDM Enabler chooses whether to provide security mechanisms to protect data communicated between SCIDM Client and CIM against any unauthorized changes and modifications. A suitable mechanism for integrity protection is specified in [OMA-SEC_CF-AD].

Deployment of the SCIDM Enabler chooses whether to provide security mechanisms to avoid the disclosure of data communicated between SCIDM Client and CIM without the permission of its owner. A suitable mechanism for confidentiality protection is specified in [OMA-SEC_CF-AD].

Appendix A. Change History (Informative)

A.1 Approved Version History

Reference	Date	Description

A.2 Draft/Candidate Version <current version> History

Document Identifier	Date	Sections	Description
Draft Versions OMA-AD-SCIDM-V1_0 Draft Versions OMA-AD-SCIDM-V1_0	24 Nov 2008	5.2,5.3	Incorporates agreed CR: OMA-SEC-SCIDM-2008-0005- CR_Arch_Diagram_and_Definition.doc
	18 Dec 2008	Appendix B	Incorporates agreed CR: OMA-SEC-SCIDM-2008-0012-CR_flows_for_IDE2.doc
	18 Dec 2008	5.3.3.1.2	Incorporates agreed CR: OMA-SEC-SCIDM-2008-0011R01- CR_offine_identification.doc
	18 Dec 2008	Appendix B	Incorporates agreed CR: OMA-SEC-SCIDM-2008-0010-CR_flows_for_REG.doc
	18 Dec 2008	Appendix B	Incorporates agreed CR: OMA-SEC-SCIDM-2008-0009-CR_flows_for_IDE.doc
	18 Dec 2008	2.2,5.4	Incorporates agreed CR: OMA-SEC-SCIDM-2008-0007- CR_AD_Security_Consideration.doc
	10 Feb. 2009	5.2	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0001-CR_Arch_Diagram_Change.doc
	10 Feb. 2009	5.3	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0002- CR_Component_Interface_Def_Changes.doc Mainly changed "component" to "function", added 5.3.3.2.1 and 5.3.3.2.2
	10 Feb. 2009	3.2, 3.3	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0005R01-CR_Ref_Abbr_Update.doc
	10 Feb. 2009	5.3.3.1.4	Incorporates agreed CR: OMA-SEC-SCIDM-2008-0008R01- CR_Interoperation_between_CIMs.doc
	10 Feb. 2009	1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0006R01-CR_Scope_Clarification.doc
	10 Feb. 2009	5.3.1.1 5.3.1.4.1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0007R01- CR_Iden_Repurposed_Compound_Partial_Content.doc
	10 Feb. 2009	5.3.1.1.1 5.3.3.1.3	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0008R01- CR_Trust_of_Registration_Info.doc
	10 Feb. 2009	4.1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0009-CR_Phase_Planning.doc
	10 Feb. 2009	B.4	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0010R01- CR_Flows_for_CIM_interoperation.doc

Document Identifier	Date	Sections	Description
	14 March 2009	5.2,5.3	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0018R01-CR_Interface_Exposed_by_Iden_Client.doc
	14 March 2009	5.3.1.1.2,5.3.1.4.1,5.3.1.4.2	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0017-CR_Section_5.3.1.1.2.doc
	14 March 2009	5.3.1.3,5.3.1.4	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0016-CR_Section_5.3.1.4.doc
	14 March 2009	5.3.3.1,	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0015-CR_Section_5.3.3.1.doc
	14 March 2009	5.3.2	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0014-CR_Section_5.3.2.doc
	14 March 2009	5.3.1.1.1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0013-CR_Section_5.3.1.1.1.doc
	14 March 2009	5.2	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0012R01-CR_Section_5.2.doc
	14 March 2009	3.2	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0011-CR_Section_3.2.doc
	24 March 2009	5.3.4.1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0027-CR_Client_Online_Offline_Clarification.doc
	24 March 2009	3.2,5.3.1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0026-CR_Def_update.doc
	24 March 2009	2.1,2.2,5.1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0024R01-CR_Dependencies.doc
	24 March 2009	5.3.3,	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0023-CR_Interface_Label_Consistency.doc
	01 May 2009	5.2,5.3.1,5.3.2,5.3.3.1.1,5.3.3.1.2,5.3.3.1.3,5.3.3.1.4,5.3.3.2	Incorporates agreed CR : OMA-SEC-SCIDM-2009-0034R01-CR_Arch_Scope.doc
	01 May 2009	5.3.1.4	Incorporates agreed CR : OMA-SEC-SCIDM-2009-0036-CR_Section_5.3.1.4.doc
	01 May 2009	2.1,2.2	Incorporates agreed CR : OMA-SEC-SCIDM-2009-0037-CR_Section_2.doc
	01 May 2009	B.1	Incorporates agreed CR : OMA-SEC-SCIDM-2009-0043-CR_A042.doc
	01 May 2009	B.2.1	Incorporates agreed CR : OMA-SEC-SCIDM-2009-0045-CR_A043.doc
	16 May 2009	5.3.1.1.1,5.3.3.1.3,5.4.1,5.4.2,5.4	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0035R02-CR_Arch_Security_Issues.doc
	16 May 2009	5.3.1.1.1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0038R01-CR_Registration_Principal_Differentiation.doc
	16 May 2009	2.1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0039R01-CR_OSE_Conformance.doc
	16 May 2009	All	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0044R01-CR_Misc_Changes_to_Normative_Sections.doc
	18 May 2009	5.3.1.1.1	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0040R01-CR_A019_A020.doc

Document Identifier	Date	Sections	Description
	18 May 2009	5.3.1.1.2	Incorporates agreed CR: OMA-SEC-SCIDM-2009-0042R01-CR_A021.doc
	18 May 2009	4	Incorporates agreed CR: OMA-ARC-SCIDM-2009-0048-CR_Section4_change.doc
	19 May 2009	B.1	Incorporates agreed CR: OMA-ARC-SCIDM-2009-0041R01-CR_A041.doc
	19 May 2009	5.3.1.1.1,5.3.1.1.2	Incorporates agreed CR: OMA-ARC-SCIDM-2009-0053-CR_Misc_Changes.doc
	26 May 2009	5.3.3.1.5	Incorporates agreed CR: OMA-ARC-SCIDM-2009-0054-CR_Remove_SCIDM4_Interface.doc
	25 June 2009	5.3.1.1, 5.3.1.1.4, 5.4	Incorporates agreed CR: OMA-ARC-SCIDM-2009-0057-CR_Fix_SCIDM_AD_per_remaining_issues.doc
Candidate Version: OMA-AD-SCIDM-V1_0	28 Jul 2009	All	Status changed to Candidate by TP: OMA-TP-2009-0324-INP_SCIDM_V1_0_AD_for_Candidate_approval

Appendix B. Flows (informative)

B.1 Flow for Content Registration

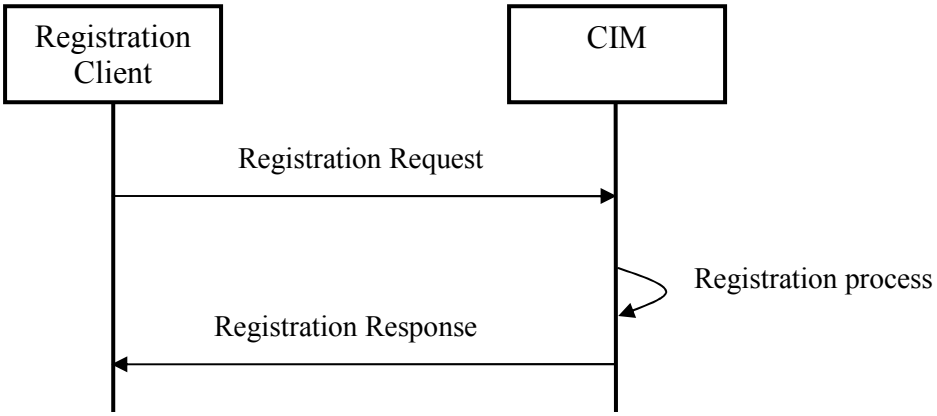

Figure 2: Flow for Content Registration

This flow is initiated by the Registration Client. The Registration Client MAY be operated by different entities depending on the application scenarios, e.g., content provider in copyright verification application, network operator or normal user in anti-virus and spam filtering applications.

1. The Registration Client sends registration request to the CIM.

   The request contains at least the following information:

   1. The content or the content fingerprint.

   The following information MAY also be included in the request message:

   a. The type of the application scenario in which the content is to be managed. This information is used to indicate the objective of registration. When the Registration Client and the CIM can maintain a consistent view on the type of application scenario through others means, this parameter needs not be present.

   b. The fingerprinting algorithm ID if the request contains the content fingerprint instead of the content.

   c. The content type and format.

   d. The Registration Client's identifier. In some application, this is not needed, e.g., in the spam reporting and registration from ordinary users.

   e. Metadata specific to the application scenario, such as the information about the copyright owner and copyright announcement in copyright verification applications.

   f. Credential information for increasing the assurance of the metadata. In registration of user generated content, this information MAY be provided by some proxy agent, such as content posting sites, and is useful for by the CIM and others to establish some trust to the ownership claim and other associated metadata.

   g. Supplemental description, e.g., the reason for the registration.

2. Upon receiving the registration request, the CIM starts the registration process.

   a. If credential information is provided, the CIM SHALL verify the credential information. If it is valid, do the following.

   b. The CIM MAY first check if the content has been registered before. If not, do the following steps.

   c. Allocate ID for the content, extract fingerprint based on the application type and content type/format if no fingerprint is included in the request.

d. Record the means used to verify the trustworthiness of the metadata provided.

e. OPTIONALLY, the CIM generates the Content ID Certificate.

f. Finally, the CIM stores relevant information in the database.

3. The CIM returns the registration response to the Registration Client.

   The response contains the result of the registration process, and optionally the Content ID Certificate.

B.2 SCIDM-2 Flow

This section describes the content identification and query flow between the Identification Client and the CIM. In identification and query, the Identification Client requests the CIM to check if a content item falls into a scope that some kind of management action SHOULD be applied to it. This is done by matching the content item to the one registered in the CIM and determine what to do based on the metadata of the registered content item.

The match in the identification process MAY be done through multiple mechanisms, including content ID, metadata (e.g., MD5 hash value), identification history, digital watermark, and content fingerprint, among which, the content fingerprint mechanism is the most secure one.

For efficiency, it is not necessary that every identification process is done by the content fingerprint mechanism. What mechanism to choose depends on the load of the system, demand on efficiency and required security level.

B.2.1 Flow for Content Identification and Query

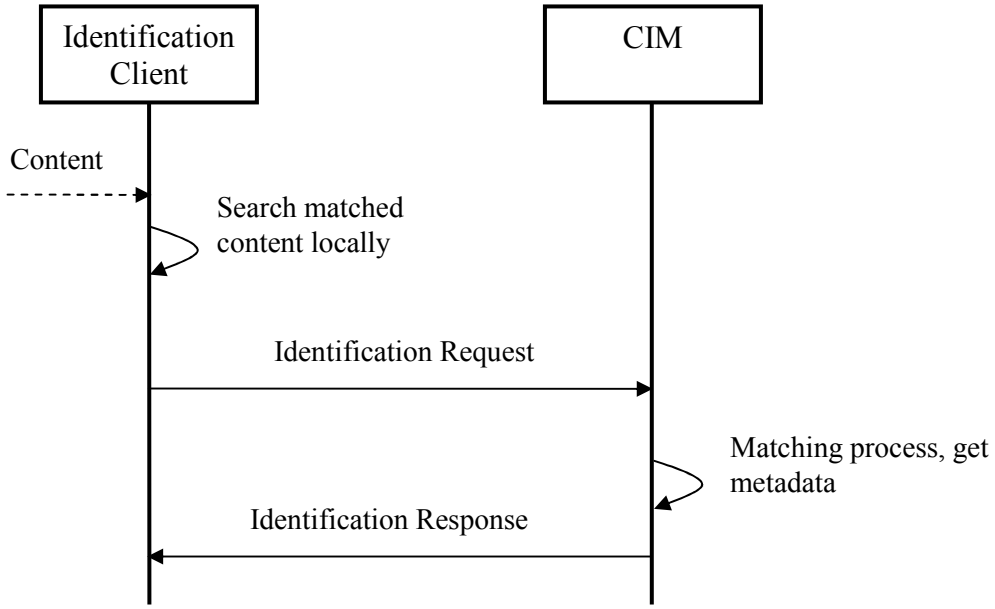

Figure 3: Content Identification Flow

This flow is initiated by the Identification Client. The Identification Client usually starts the identification and query process upon receiving a content identification request from the Content Management Entity (CME), and a content item is passed to the client from the CME. For more details about the CME and its relation with the client, please see Section 5.3.1.4.

1. The Identification Client first searches for the matched content locally.
   In Identification Client, there MAY be some content registration information, which MAY be aquired from former synchronization from the CIM. If no match is found, the Identification Client does the following steps.

2. The Identification Client sends identification request to the CIM.
   The request MAY contain the following information:

   a. The type of application scenario in which the content is to be managed.
   b. Client's preference of content identification mechanism
   c. Basic content information, such as ID, name, type, format, etc.
   d. Content delivery information, such as the source address and destination address.
   e. If the client's preferred identification mechanism is by fingerprint, then the content or the content fingerprint SHALL be included. The watermark mechanism is considered one specific kind of fingerprint. As to watermark mechanism, the fingerprint refers to the watermark embedded in the content.
   f. The fingerprinting algorithm ID if the request contains the content fingerprint instead of the content.
   g. Information specific to the application scenario.

3. The CIM searches for the matched content.
   The CIM SHALL select the appropriate content identification mechanism based on the application type, content type and client's preference. If matched content is found, it retrieves the necessary metadata from the Content Metadata Database.

4. The CIM returns the response to the Identification Client.
   If the identification and query process succeeds in Step 3, the response contains the following information:

   a. The identification result code
   b. Requested content metedata for the specified application
   c. In some cases, Content ID Certificate.

   If the identification and query process fails in Step 3, the response contains the following information:

   a. The identification result code
   b. OPTIONALLY the suggested mechanisms for identification by CIM. If the client's preferred mechanism in the request is not fingerprint mechanism and no content or content fingerprint is included in the request, the CIM SHOULD give suggestion in the response that the identification MAY succeed via the fingerprint mechanism.
      The Identification Client MAY initiate a new identification process with a different identification mechanism. The choice of new mechanism is based on the CIM's suggestion, client's load, or demand on the efficiency.

B.3 SCIDM-3 Flows

B.3.1 Flow for Identification Client Initiated Synchronization

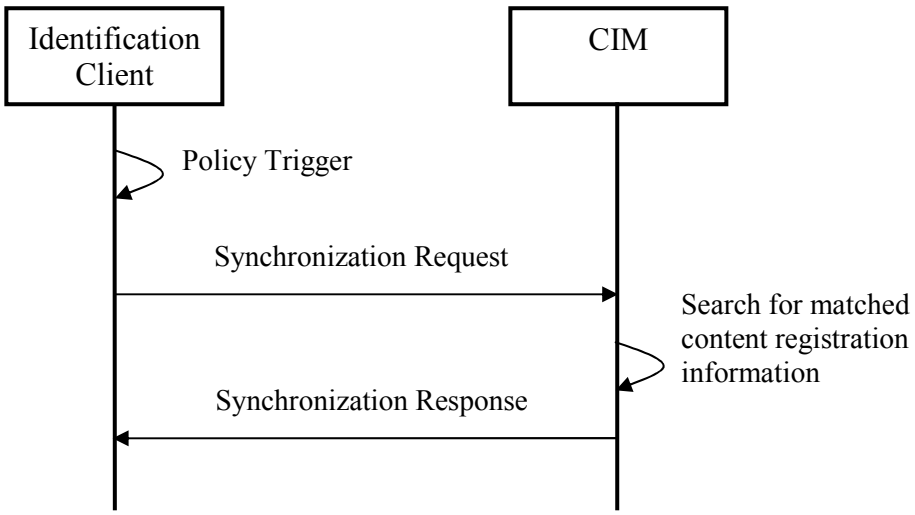

Figure 4: Flow for Identification Client Initiated Synchronization

This flow is initiated by the Identification Client when the pre-configured policy on the Identification Client has been triggered.

1. The pre-configured policy is triggered when the pre-defined condition is met. See section 5.3.3.1.2 for the possible scenarios that the policy is defined.
2. The Identification Client sends a synchronization request message to CIM to retrieve content registration information.

   The request message contains the same information as the identification request in the SCIDM-2 interface except for

   a label indicating that this is a synchronizaiton request initiated by the Identifiation Client.

   The request message MAY contain information of multiple content items to request synchronization of multiple content registration information.
3. CIM searches for the matched content registration information. The search process is the same as in the identification process for the identification request of SCIDM-2.
4. CIM provides a response message to the Identification Client. This message contains:
   a. the content ID certificate for each content item in the request.

B.3.2 Flow for CIM Initiated Synchronization

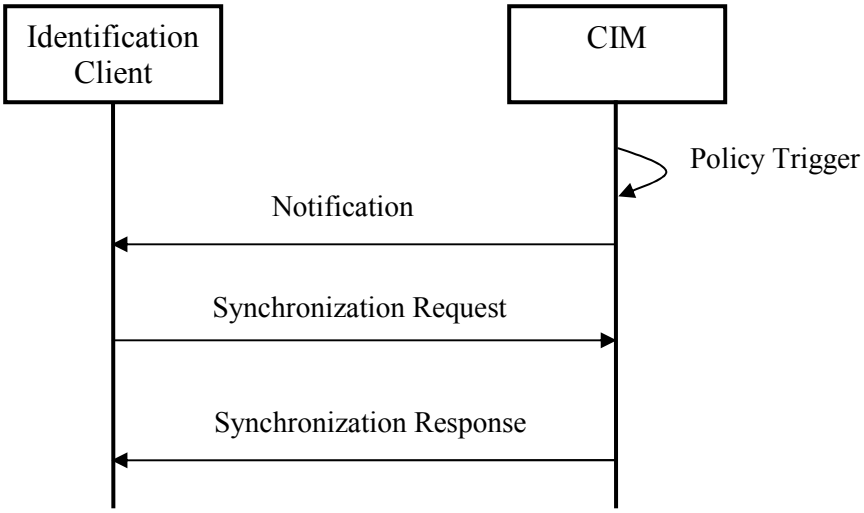

Figure 5. Flow for CIM initiated synchronization

This flow is initiated by the CIM when the pre-configured policy has been triggered in the CIM.

1. The pre-configured policy is triggered when the pre-defined condition is met. See section 5.3.3.1.2 for the possible scenarios that the policy is defined.
2. CIM sends a notification to the Identification Client to inform that some content registration information need to be synchronized to the client. The notification MAY be sent via various available mechanisms, such as SMS/MMS, WAP Push, etc.
3. The Identification Client starts the synchronization process upon receiving the notificaiton. The client sends the synchronization request to the CIM. The request contains the following information:
   a. The identifier information of the notification from the CIM, such as the message id, or the session id.
   b. A label that indicates this is a synchronizaton request initiated by the CIM.
4. The CIM returns the content registration information that it intends to synchonize to the client in the form of Content ID Certificate.

B.4 Flows for CIM Interoperation

The interoperation between CIMs is used to achieve content identification when there is no registration information of the to-be-identified content item in the local CIM. Two modes of interoperation are RECOMMENDED, as discussed below.

B.4.1 Interoperation in the Central CIM Mode

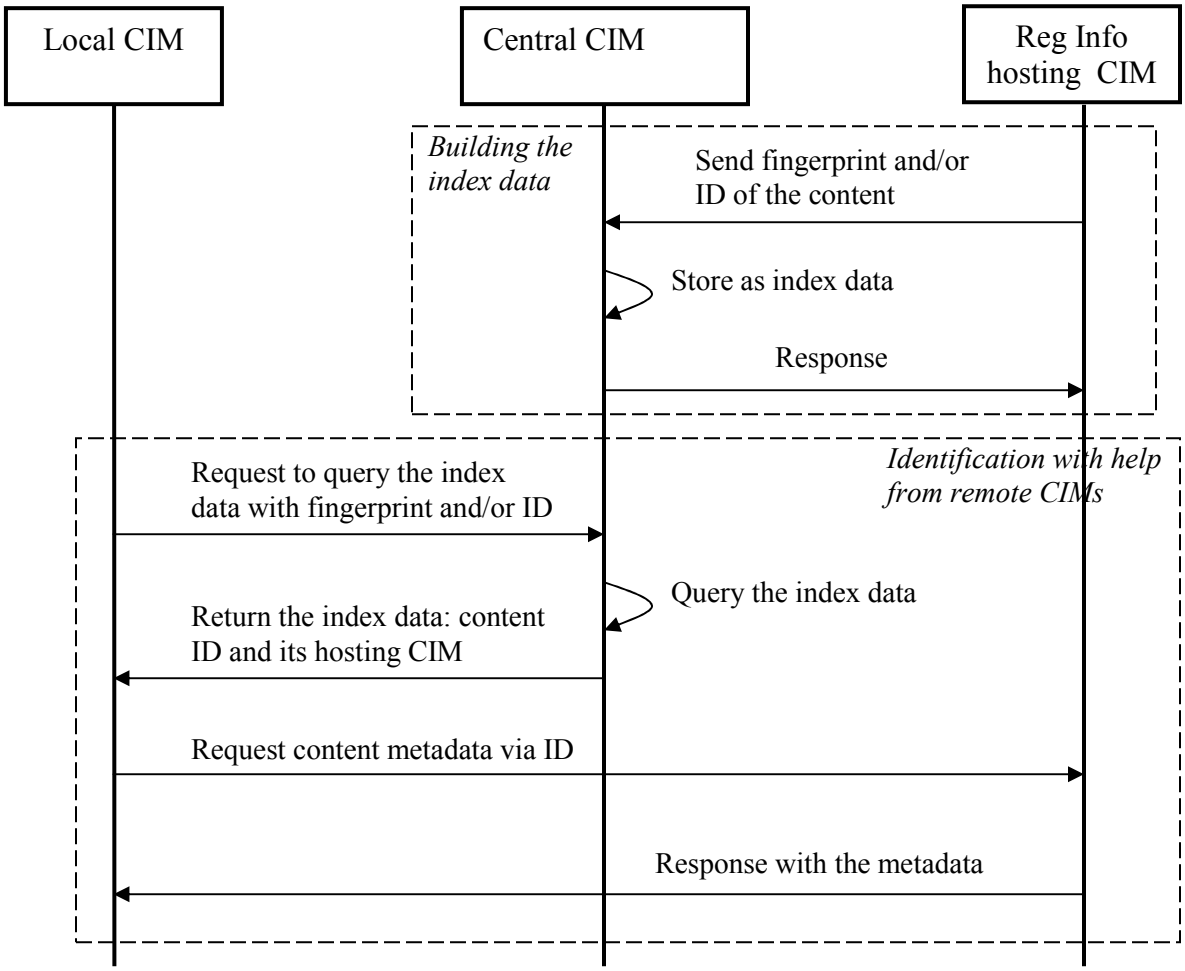

Figure 6: Flow for central CIM mode

In this mode, there is a Central CIM which stores the index data. This flow has two phases; the first is to build the index data, and the second is to identify the content item with the help of remote CIMs, i.e., the Central CIM and the registration information hosting CIM.

In the first phase, i.e., building the index data,

1. After a CIM (the registration information hosting CIM) completes the registration of some content item, it sends the fingerprint and/or ID of the content item to the Central CIM.

2. The Central CIM stores the mapping between the fingerprint and/or ID of the content item and the ID of the registration information hosting CIM as the index data.

3. The Central CIM sends a response to the registration information hosting CIM.

Note: the SCIDM-1 interface MAY be used in the first phase.

In the second phase, i.e., content identification with the help of remote CIMs,

1. When there is no registration information of the queried content item in the local CIM, the local CIM sends a request to the Central CIM to query the index data. The request SHALL contain the fingerprint and/or the ID of the content to be identified.

2. The Central CIM queries the index data.

3. If the Central CIM succeeds in finding the corresponding index data, it returns it to the local CIM.

4. The local CIM gets the ID of the registration information hosting CIM, and sends the request to it to retrieve the metadata of the content item.

5. The registration information hosting CIM returns the requested metadata to the local CIM.

Note: the SCIDM-2 interface SHALL be used in the second phase.

B.4.2 Interoperation in the CIM List Mode

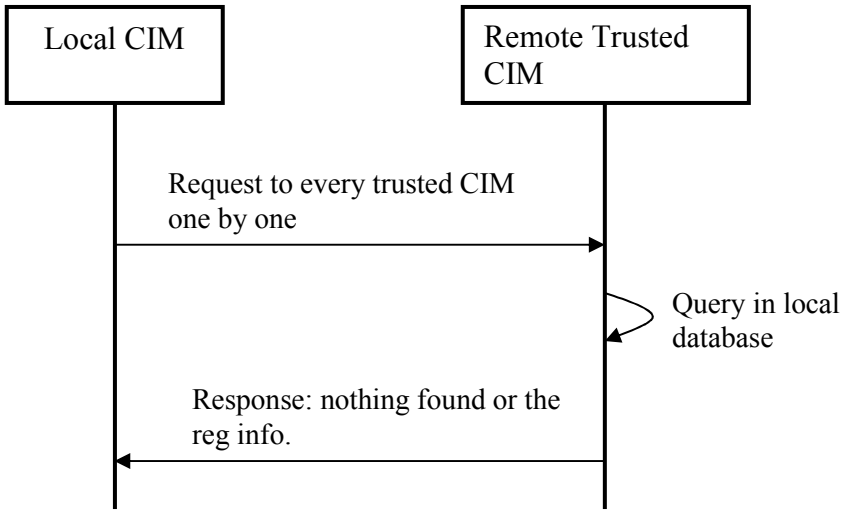

Figure 7: Flow for CIM list mode

In this mode, every CIM has a list of IDs of remote CIMs which it has a trust relation with. The flow is:

1. When there is no registration information of the queried content item in the local CIM, the local CIM sends a request to the remote CIMs that are in the list one by one. The request contains the fingerprint and/or the ID of the content item. Upon receiving the request from the local CIM, the remote CIM queries the database based on the received fingerprint and/or Content ID.
2. If the corresponding registration information is found, the remote CIM sends a response with the registration information, if not found a response is also returned.

Note: A local CIM SHOULD query all of the remote CIMs in the list in order to identify the duplicate registration scenario. The SCIDM-2 interface SHALL be used here.

Appendix C. Deployment of SCIDM Enabler (informative)

The SCIDM Enabler provides a basic content identification service and its deployment is highly dependent on its application. In this section, the following two general deployment issues are illustrated for better understanding the usage of SCIDM:

- The usage of SCIDM Enabler by applications
- The Interoperation between Local CIM and Remote CIM

C.1 Usage of the SCIDM Enabler by applications

The following figure illustrates how the applications/services can use the SCIDM Enabler.

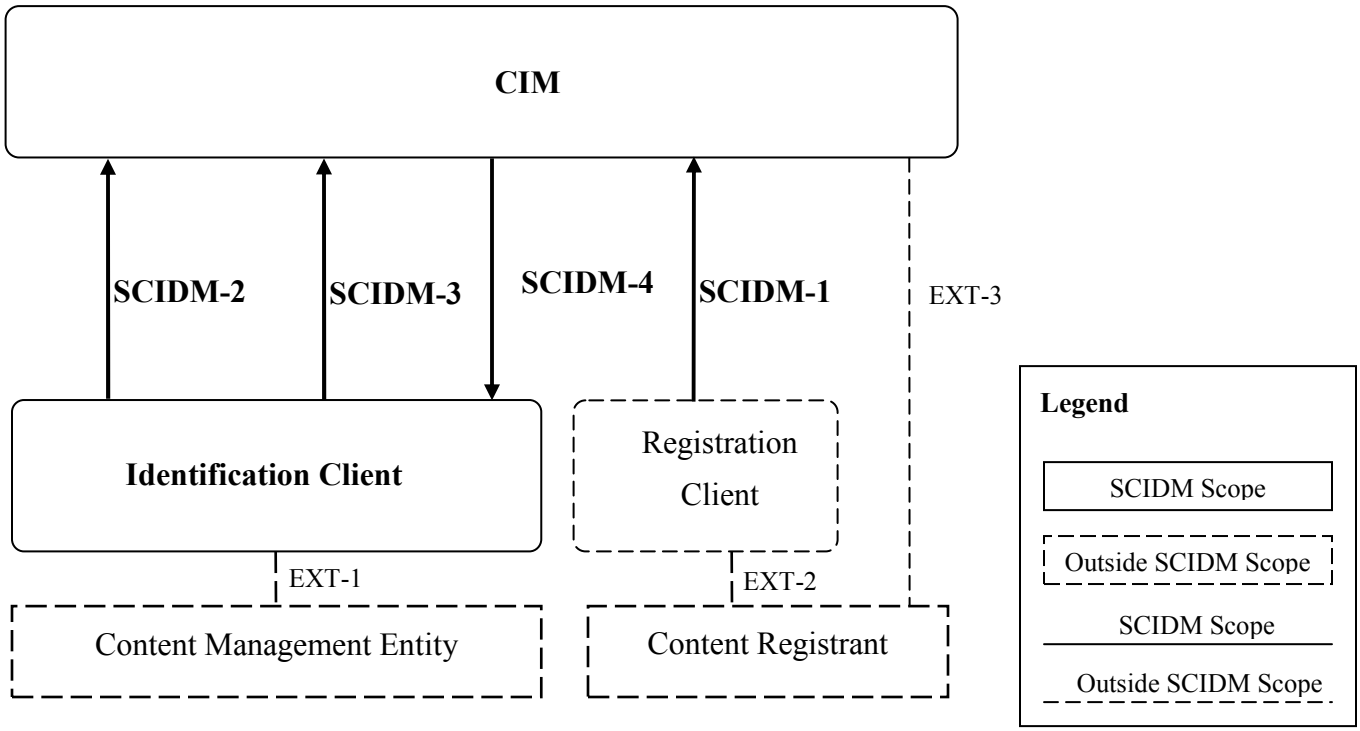

Figure 2: Interactions of the SCIDM Enabler with External Applications

The entities and interfaces that are in scope of the enabler are defined in Section 5. The external entities Content Management Entity and Content Registrant and the external interfaces EXT-1, EXT-2 and EXT-3 are not defined by the SCIDM Enabler.

The Content Registrant is the entity that provides content to CIM for registration for some purpose, such as that the content needs to be copyright protected, or to be filtered, etc. It can be the Content Provider, the SCIDM Service Provider, Network Operator, or normal user, depending on the application.

The Content Management Entity performs some type of content management, such as monitoring of infringement content, spam filtering, etc. The Content Management Entity may reside in network side, user side (i.e. user device), which depend on the application. And generally the Identification Client resides in the Content Management Entity and act as its logical function for content identificaiton. If the Content Management Entity needs to identify if it should take action on a content item, it uses the Identification Client to interact with CIM for content identification. If it successfully gets result from the Identification Client, and in the result some metadata is included, it should take the management action based on the metadata.

EXT-1 is used by the Content Management Entity to send content identification request to the Identification Client and receive the response. EXT-2 is used by the Content Registrant to send content registration request to the Registration Client and receive the response. EXT-3 is used for registration mechanisms other than that supported by SCIDM-1, e.g., registration via CIM Portal, via ftp, via email, etc. The definition of these interfaces is out of scope of the SCIDM Enabler.

C.2 Interoperation between Local CIM and Remote CIM

When there is no registration information of a content item in one CIM, the CIM can contact other CIMs to accomplish the content identification task. Although this is not mandatory, yet in situations where the CIM has some trust relationship with other CIMs, that is, they belong to a CIM trust group, it SHOULD contact the relevant CIMs for content identification. The trust relationship MAY be established by some means, such as:

a. The CIMs are administrated by the same operator.

b. The CIMs have a federation relation set up by operators in the form of an agreement.

Two types of interoperation mode between the CIMs are RECOMMENDED here: the Central CIM mode and the CIM List mode. Which mode is used depends on the operator of the CIM.

Note: It is recognized that a race condition MAY occur when identical content is registered with multiple CIMs simultaneously. A race condition MAY also occur when a new CIM joins a trust group carrying identical content. Therefore a mechanism is needed to reconcile multiple registrations for the same content. The method of reconciliation is out of scope.

C.2.1 Interoperation in the Central CIM mode

In this mode, there is a Central CIM which stores the index data, i.e., the mapping between the ID of content registration information and its hosting CIM. The ID of content registration information contains the fingerprint value and the Content ID assigned by the hosting CIM of the content.

After a content item is registered successfully to a CIM, which is the hosting CIM of the registered information, the CIM SHOULD send the fingerprint and/or the ID of the content to the Central CIM for creating the index data. The CIM needs not send all the registration information to the Central CIM. SCIDM-1 MAY be used for this purpose, i.e. the hosting CIM register the content with limited information (content fingerprint, its hosting CIM's ID) to the Central CIM.

When a local CIM needs to identify a content item that is not registered locally, it SHOULD send the fingerprint and/or the ID of the content to the Central CIM for querying the index data. If the corresponding index data is found, the Local CIM will receive the index data from the Central CIM. SCIDM-2 MAY be used for this purpose with the Local CIM as Identification Client and the Central CIM as the Server.

Via the index data, the Local CIM can retrieve the content registration information from the hosting CIM (another Remote CIM). If no index data is found, the Local CIM MAY conclude that the content item is not registered. SCIDM-2 also MAY be used for this interoperation with the Local CIM as the Identification Client and the hosting CIM as the Identification Server.

C.2.2 Interoperation in the CIM List mode

CIM list is a list of IDs of Remote CIMs that has a trust relation with the Local CIM. When the Local CIM needs to identify a content that is not locally registered, it sends a request to every remote CIM in the list for identification. If some Remote CIM has the registration information, it SHALL return them to the CIM. If no Remote CIM returns the needed information, the CIM MAY conclude that the content item is not registered. SCIDM-2 MAY be used for this purpose with the Local CIM as the Identification Client and the Remote CIM as the Identification Server.

The CIM List is created when the CIM trust group is established. When a new CIM joins the trust group, its ID SHOULD be added to the List. Each CIM in the group SHOULD have the List. The creation and the update of the List MAY be done manually or by some technical mechanisms which is out of the scope of this enabler.

C.2.3 Cache the registration information locally

In both modes, if the registration information for a content item is found in a Remote CIM, the CIM shall record which CIM contains the registration information for later fast look-up. Otherwise, it should record that the content is not registered in the trust group. This is to avoid subsequent unnecessary interoperation with Remote CIMs.