2005 年 1 月

利用状況を把握するコンテンツ ID による認証

１．はじめに

ユビキタスネットワーク社会においては，種々のメディアのコンテンツがネットワーク中を流通するが、コンテンツの流通状況を調べることは、図 x-1 に示すように、三つの点で重要である。まず、利用状況、すなわち、あるコンテンツがどれだけネット上でダウンロードされているか、あるいは視聴されているかという統計情報を把握するマーケティングである。次に、そのコンテンツの流通履歴の把握がある。コンテンツを制作したクリエータから始まり、ディストビリュータ等種々の事業者を通ってエンドユーザまで行くが、そのエンドユーザ同士がまたＰ２Ｐ形態でコンテンツを交換するかもしれない。こうしたコンテンツが流通する履歴を把握することも、素性の確かなコンテンツを利用するという消費者保護の点から重要である。三つ目が、不正利用監視・追跡であり、正規の許諾を受けた利用の範囲内なのかそうでないのかをチェックする権利者保護である。このようなコンテンツ流通情報の収集・管理に必要な技術の一つが、コンテンツ認証である。以下では特に、不正利用監視・追跡に焦点を当て、コンテンツ ID を用いたコンテンツ認証技術がどのように使われているかについて解説する。

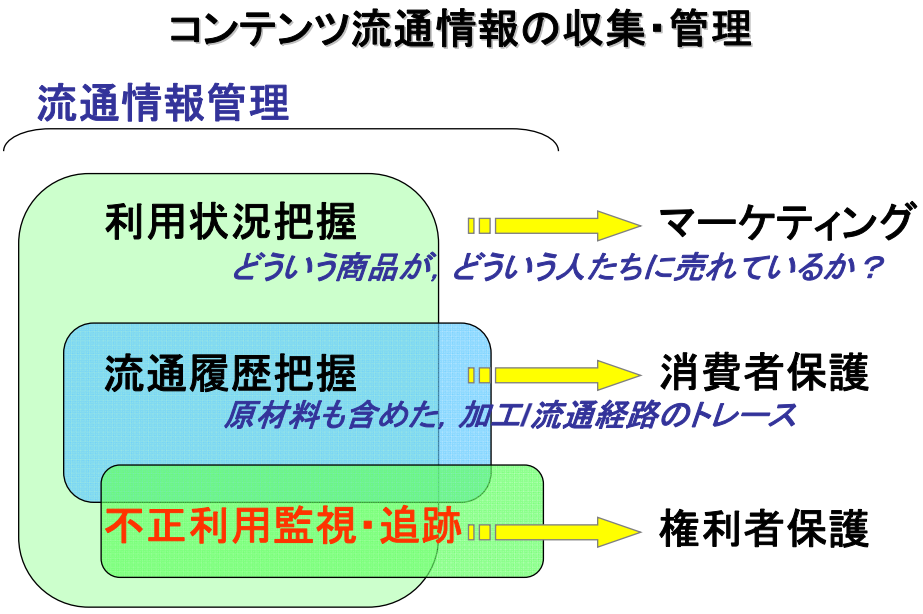

図 x-1　コンテンツ流通情報の収集・管理

２．　不正利用監視・追跡技術

コンテンツの流通情報管理のうち、特に、不正利用監視・追跡は、次に示す大まかな流

Hiroyuki YAMASHITA
cIDf

れにより実現される。[1]

まず、流通するコンテンツに関する情報を *「収集」* する。これは、どこそこにどんなコンテンツがある、といったことである。次に、収集したコンテンツ情報の中から、監視対象を *「識別」* する。これは、手当たり次第にチェックすると処理が膨大になるため、あらかじめ“くさい”ところにチェック対象を絞り込むことである。例えば、ある映画が不正に使われているらしいと分かれば、その映画を、収集したコンテンツ情報の中から識別する訳である。続いて、監視対象（この映画）コンテンツが正規に許諾を受けた利用になっているか不正利用かどうかを、チェック（*「分析」*）する。その結果、不正だと分かれば、警告とか告発といった *「アクション」* をとる。なお、収集された情報は整理して *「管理」* され、必要に応じ後々の収集・分析等に使われる。

以上示した収集・識別・分析・アクション、それぞれのフェイズについて、フェイズごとにどのような技術が使われるかについて、図 x-2 に示す典型的な例を用いて説明する。同図の例は、最も単純な、ＩＤと電子透かしを利用するものである。[2] コンテンツに付与されたユニークなＩＤを電子透かしで埋め込んでおき、そのコンテンツの流通情報、権利情報等をＩＤ管理センタのデータベースに格納しておく。電子透かしでＩＤを埋め込んだコンテンツを正規に流通させるが、流通の過程で悪意を持った利用者がこのコンテンツを無断でコピーし、例えば自分のホームページに掲載して使用しているとする。

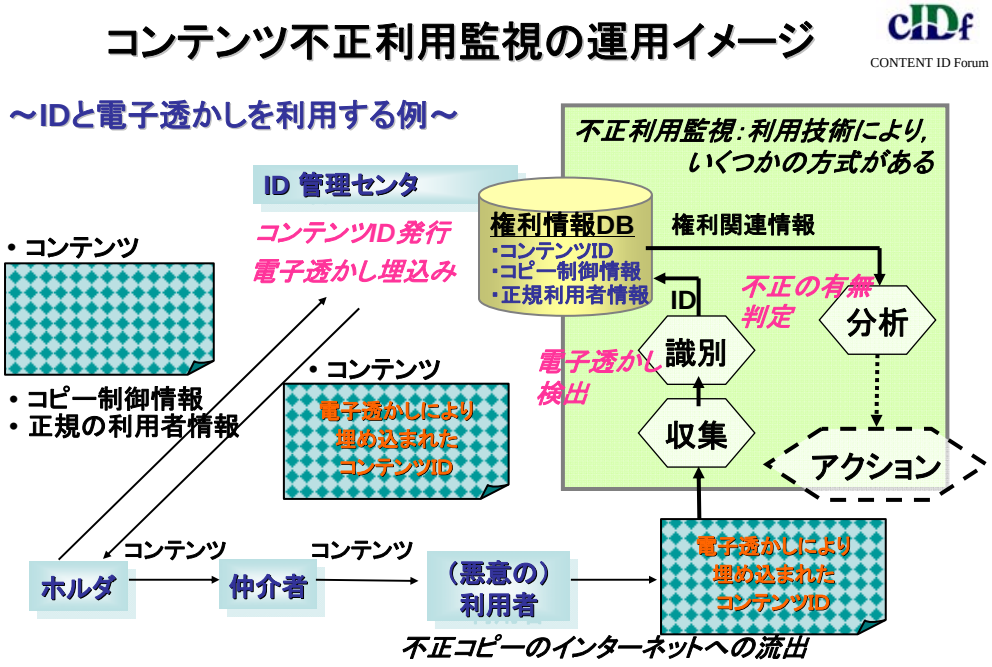

図 x-2　コンテンツ不正利用監視・追跡の一例

このような状況において、疑わしいコンテンツが見つかると、そのコンテンツが置かれ

ているホームページの情報をコンテンツ情報収集技術で収集する。収集したコンテンツには電子透かしでＩＤが埋め込まれているため、それを抽出する。そのＩＤに対応する権利情報が ID 管理センタのデータベース中にあるので、そのＩＤをキーにセンタのデータベースにアクセスし、権利関連情報を取り出して参照する。分析では、この人のホームページにこのコンテンツが掲載されることが許諾されているかどうかをチェックする。不正利用が確認されると、相応のアクションを起こすことになる。

以降では、不正利用監視・追跡の流れにおける各フェイズで使用される技術について説明する。

２．１　　コンテンツ情報の収集

コンテンツ情報の収集方式には、どこで収集するか、どういう方法で収集するかという観点から、幾つかの型がある。その代表が図 x-3 に示す３つである。

①探索ロボット型

②ネットワークノード型

③端末内蔵型（利用者協力型）

図に示すうち、現在では、探索ロボット型が最も多く使われている。ネットワークノード型は、個人情報保護の点でいくつかの問題が指摘されている。端末内蔵型（利用者協力型）は、利用者にチェック用プログラムをインストールしてもらう必要があることが、広い普及に際しての一つの障壁となっている。

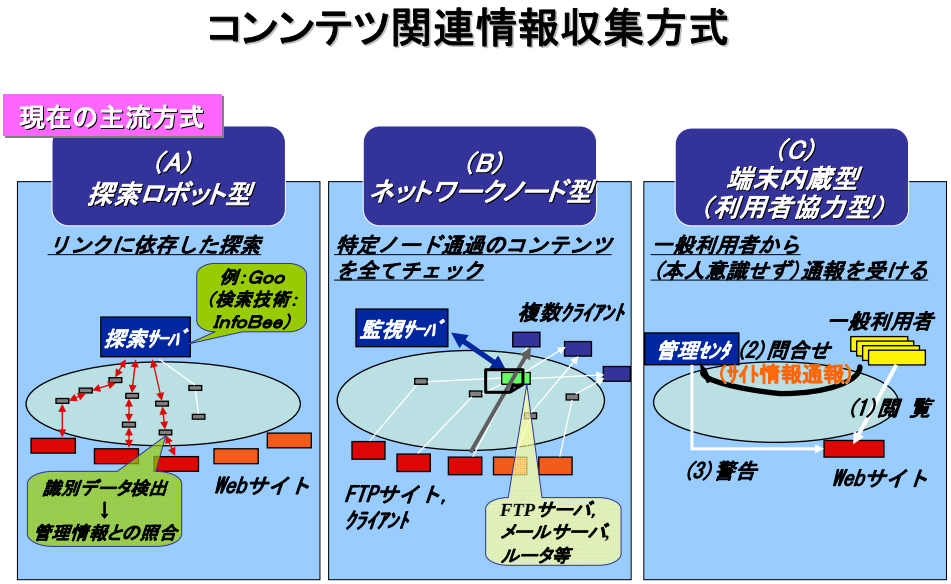

図 x-3　コンテンツ情報の収集方式

① 探索ロボット型

探索ロボット型では、インターネットの検索エンジンで使われている技術をそのまま使い、探索サーバがインターネットの各ホームページにアクセスし、コンテンツあるいはコンテンツに関する情報を収集する。その後、例えばコンテンツに埋め込まれたＩＤを抽出し、それをキーとして管理センタにアクセスして得られる著作権管理情報を収集情報と照合することにより、そのコンテンツ利用が不正かどうかをチェックする。なお、インターネットのホームページは非常に膨大な量があり、それを全て探索するには何年かかるか分からない。したがって、実際に探索ロボット型を使う場合には、通常、怪しそうなサイト群を事前に特定しておき、それらを対象に探索をかける。

② ネットワークノード型

Ｐ２Ｐ形態をはじめほとんどに当てはまることは、インターネットでコンテンツをやりとりする際には、基本的にはネットワークノード（ファイルサーバと、メールサーバ、あるいはルータ等のネットワーク装置）を経由して、個人から個人、あるいは個人からサーバ、サーバから個人へコンテンツが移動することである。その中の装置にプローブを挿入し、そこを通過するコンテンツをコピーする。言い換えれば、ちょっとわき道に入れて、そのコンテンツをチェックするというのが、ネットワークノード型である。おおまかに言えば、ネットワークの中で、コンテンツ情報をトラックしてそのＩＤをチェックし、データベース情報と比較するというものである。コンテンツ情報をトラックした後は、ＩＤ付きであれば、ＩＤをもとにデータベース中の権利情報と照合する点では、ロボット探索型をはじめとする他の型と同じである。どこでコンテンツを捕捉するかが、それぞれの型により異なる。

③ 端末内蔵型（利用者協力型）

端末内蔵型では、例えば、一般ユーザがインターネットをブラウジングする際に、ブラウザソフトのバックグラウンドで、ユーザにほとんど意識させずに、コンテンツ情報収集用のプログラムが動作する。そのプログラムは、ユーザがブラウズするホームページに掲載されているコンテンツの情報をモニタし、不正かどうかを分析する。[3] この型は、利用者がたまたま見ているホームページ上のコンテンツをチェックするということから、利用者協力型とも言われる。このような機能の実現方法としては、例えば、ブラウザソフトのプラグインのようなものが考えられる。ただし、そういうソフトを利用者が自分の端末にインストールすることを許容してくれないと、うまくワークしない。

○ 各方式の比較

コンテンツ情報収集方式には、以上説明したように３つの代表的な型があるが、それぞ

れ一長一短がある。それを比較して簡単にまとめたのが、表 x-1 である。比較の観点は、探索範囲、すなわち、どの程度のコンテンツをカバーできるかという点、ネットワークにかける負荷、及び、それがどれぐらい簡単にできるかという点である。表に示すように、各方式には一長一短があるため、実際には、複数の方式を組み合わせたり選択したりすることが必要になると考えられる。

表 x-1　コンテンツ情報収集方式の比較

各情報収集方式の比較

方式	探索範囲	ネットワーク負荷	簡便さ
(1) 探索ロボット型	○ インターネットのオープンサイト全てが対象	× 探索サーバに接続するネットワークトラフィック大	○ Web 検索エンジン等利用で容易
(2) ネットワークノード型	○ 監視対象ノード通過コンテンツの全てが対象 P2P 流通にも適用可能	△ 特定ノードの負荷大	× メール/プロキシ等のサーバに細工が必要 プライバシ問題あり
(3) 端末内蔵（利用者協力）型	△ 監視対象端末内で扱うコンテンツは全て可能（クローズドサイトも可）	○ ネットワークの負荷はなし	△ 利用者の端末への組込み手段（インセンティブなど）が必要

評価基準) ○:ほぼ完全(容易)に満足, △:条件付きで満足, ×:実現は困難

２．２　対象コンテンツの識別（ＩＤ）

コンテンツを効率的に識別するためには、コンテンツにユニークな識別番号（ＩＤ）を付与しておく必要がある。また、付与したＩＤがコンテンツと強固に結びつけられていないと意味がない。したがって、ＩＤの付与方法と、ＩＤのコンテンツへのバインド方法とが、ここでの技術的な問題である。

ＩＤの付与方法としては、世界中で重複のない標準IDを使えばよい。このようなＩＤとしては、コンテンツＩＤフォーラム（cIDf）の提唱する“コンテンツＩＤ”がある。[2]

コンテンツへのＩＤのバインドには幾つか方法があり、そのうち最も単純なものは、ファイル名にＩＤを埋め込むものである。しかし、ファイル名は容易に変えられてしまうため、この方法は採用し難い。強固に結びつけるためには、コンテンツそのものの中に何らかの形、それもできるだけ変えられない、あるいははがせないような形でＩＤを埋め込む必要がる。そのような方法としては、コンテンツのファイルの「ヘッダ域」（表示されない部分）に設定する方法や、「電子透かし（Watermarking)」がある。あるいは、コンテンツ

も生体と同じように固有の情報を持っていることに着目し、番号ではないものの、その固有情報をＩＤとして使う「電子指紋（Fingerprinting)」がある。ここでは、これら３方法について簡単に説明する。

① ヘッダ域への ID 格納（図 x-4）

デジタル化されたコンテンツのデータ構造は、信号の領域を挟んで、ヘッダ域とフッタ域というように大きく分けられる。信号データの領域はコンテンツを表示するために使われるが、ヘッダ域はファイルの作成日等の関連情報を格納しておくために使われ、表示には直接は使われない。原コンテンツをデジタル化して符号化した後に、ヘッダの領域にＩＤやその他のメタデータを設定しておく。コンテンツが端末側に転送されたときに、信号データの領域はコンテンツの表示に使われ、ヘッダ域はそれを解析するプログラムによりその内容が読まれ、ＩＤ等が確認される。この場合、コンテンツのヘッダ域のためには、静止画のＪＰＥＧ形式の場合にはこのフィールドが使えるとか、ＭＰＥＧ形式の場合にはこういう形で使えるとかいうことが、コンテンツの符号化方式ごとに決まっており、それに対応した形で情報を埋め込む必要がある。

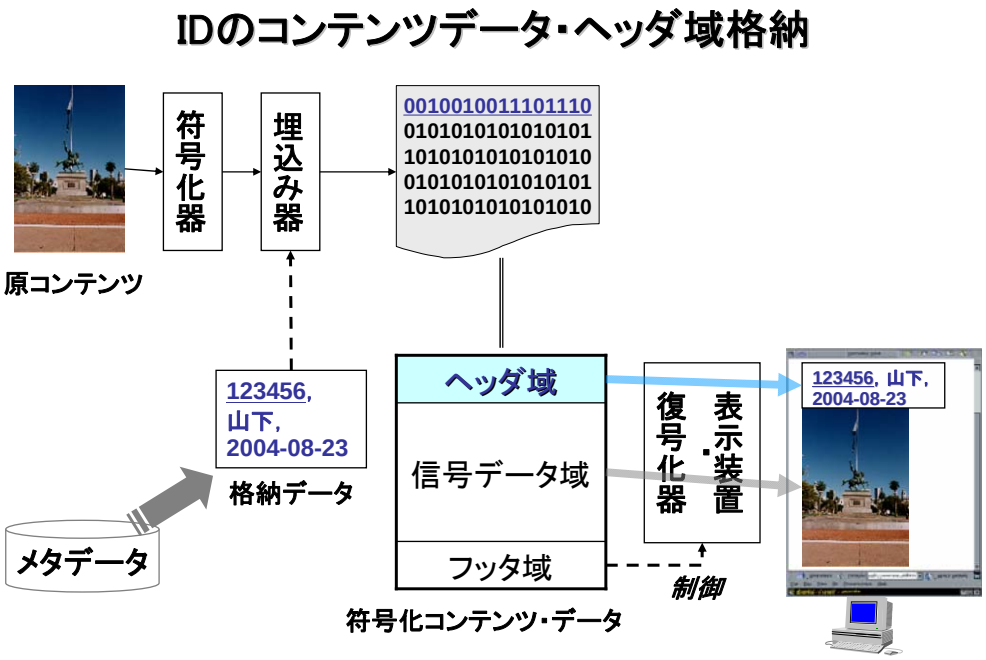

図-4　コンテンツデータのヘッダ域へのＩＤ格納例

② 電子透かし（図 x-5）

電子透かしは、人間の目には見えないような形で、コンテンツの信号域のデータを一部、ある一定のアルゴリズムに基づいて変更することによりＩＤを埋め込む方式である。もと

のコンテンツを符号化すると、ビット列が得られる。このビット列に対して、どの部分をどういうふうに変えれば１と０の情報を埋め込むことができるかというアルゴリズムがあり、それに基づいて信号域のビット列の一部を書き換える。こうして変更されたコンテンツデータが流通し、端末に送られたとする。端末での再生時には信号データ域を表示するが、流通前に行った変更は、人間の目にはほとんど見えない。電子透かしで埋め込んだ時と逆のアルゴリズムで検出することにより、埋め込まれたＩＤ情報を抽出することができる。

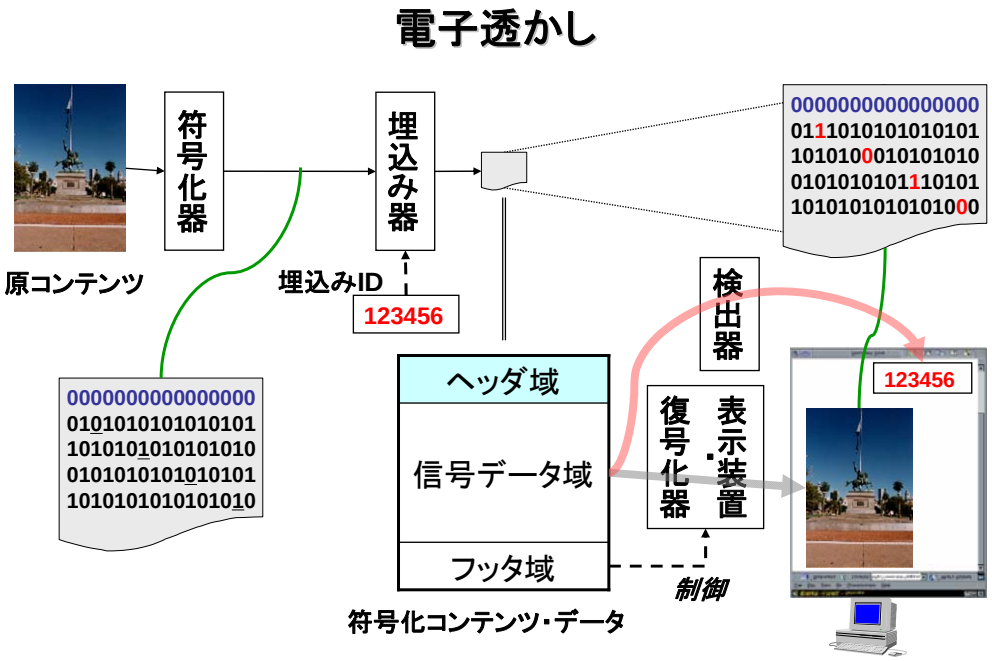

図 x-5　電子透かしによるＩＤ埋め込みの例

③ 電子指紋（図 x-6）

電子指紋はＩＤを埋め込むというものではなく、コンテンツそのものは生体と同じようにそれぞれ特徴を持っていることを利用して、その特徴を特定のアルゴリズムに基づいて抽出するものである。コンテンツを符号化した後、特徴量を抽出する関数を実行する特徴量抽出器により抽出した特徴量データをデータベースに保存しておく。コンテンツを流通させた後、ある端末に送られたものとする。その端末では、同じ関数を実行する特徴量抽出器により特徴量を抽出し、そのデータとデータベースに格納されているあらかじめ抽出したデータとを比較し、一致／不一致のチェックを行う。ここで、ＩＤに電子指紋を対応づけるには、電子指紋とＩＤとの対応表をデータベースに入れておけばよい。

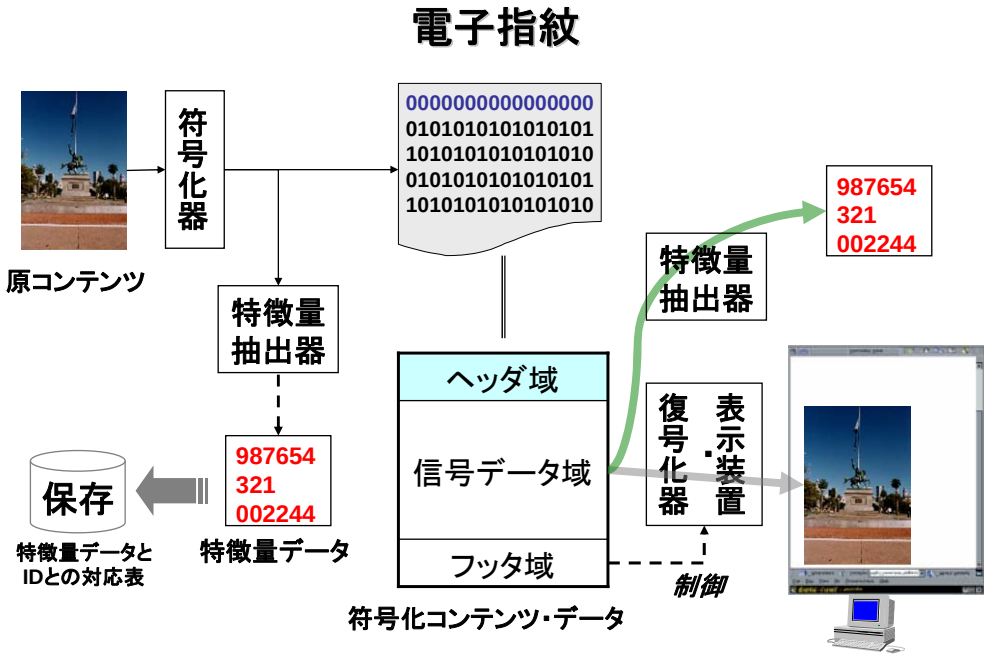

図 x-6　電子指紋によるＩＤ埋め込みの例

○方式比較

以上説明した、ヘッダ域、電子透かし、電子指紋の３つのコンテンツ識別方式についての比較を、表x-2に示す。

表 x-2　コンテンツ識別方式の比較

各識別方式の比較

比較項目	ヘッダ域	電子透かし	電子指紋
コンテンツへの事前処理	必要	必要	なし DB作成要
埋込み場所	データ構造上信号データ以外の箇所	信号データそのもの	なし
埋込み可能データ量	大	埋込み量が多いと信号データに影響あり	埋め込まない
結合の強固性	弱	中	強
信頼度	署名がある場合高 署名が無い場合低	検出率とのトレードオフ	検出率とのトレードオフ ユニーク性は関数依存
検出データ	ID	ID	特徴量
検出コスト	低	中	高
検出時間	短	長	長
印刷物から検出	不可	可	可

２．３　コンテンツ利用の分析

監視対象のコンテンツが識別された後、それが不正に利用されているか否かをチェックする方法について説明する。このチェック、すなわち、コンテンツ利用の分析には、対象コンテンツと共に収集される情報と、対象コンテンツに関連してあらかじめ管理されている情報とを用いる。

(1) 収集される情報

まず、収集される情報については、コンテンツ情報収集方式によって異なり、以下の通りである。

・探索ロボット型――ホスト（サーバ）名、URL、ファイル名称/タイプ/サイズ、収集時刻、検出複製数、ID 検知の有無/ID 値、等

・ネットワークノード型――発信元アドレス、受信先アドレス群、ファイル名称/タイプ/サイズ、収集時刻、検出複製数（累積）、ID 検知の有無/ID 値、等

・端末内蔵型（利用者協力型）――ホスト（サーバ）名、URL、ファイル名称/タイプ/サイズ、収集時刻、検出複製数、ID 検知の有無/ID 値、等

上記の他に、ＩＳＰ（プロバイダ）が保有する、次の情報を、必要に応じて使用する。

――対象コンテンツの保有者（アップロード者）情報、アップロード時刻、等

(2) ID を付与する側が管理している情報

次に、コンテンツを流通させる前に取得しデータベース等で管理している、次の権利情

報やＩＤ情報を使用する。

・権利許諾情報――利用者のホスト（サーバ）名/URL/アドレス、利用方法（Web 掲載等）、利用期間、ファイル名称/タイプ（MPEG-1/2/4 等）/サイズ、複製数、等

・ID 情報――埋め込み方法、ID 値、等

(3) 不正利用判定ロジック

上述の収集情報と管理情報とを用いて、コンテンツ不正利用の有無を判定する。その判定論理の例を、以下に示す。

・当該ホスト（サーバ）へのアップロードは許諾されているか？
・当該受信先への配布は許諾されているか？
・Web への掲載利用は許諾されているか？
・許諾された使用期間を超えていないか？
・許諾されたファイルタイプか否か？
・元のファイルは改変されていないか？（サイズの違い）
・ヘッダ域のメタデータは改竄されていないか？（署名のチェック）
・許諾された複製数を超えていないか？

２．４　コンテンツ不正利用発見時のアクション

コンテンツ利用の分析により、不正利用の疑いがあると判定された場合に起こすアクションとしては、

・ 発見と報告
・ 被疑者の特定

がある。[4] これらに関しては、個人情報や基本的人権の保護といった課題もあり、技術というよりむしろ、社会システムとして整備する必要がある。

３．　おわりに

ユビキタスネットワークにおけるコンテンツの利用状況を把握するために、コンテンツIDを利用してコンテンツ認証を行う技術について、特に、不正利用監視・追跡技術を中心に述べた。ここで、キーとなるのは、IDのコンテンツへのバインド方法である。また、コンテンツの利用状況を網羅的に把握することは、技術的には可能であっても、経済的に妥当な範囲内で実現する必要があり、運用上の工夫が求められる。

【参考文献】

[1] “コンテンツ不正利用等監視・追跡技術の利用とその法的課題に関する調査研究報告書”，DCAJ 15-CC-L，（財）デジタルコンテンツ協会(DCAj)，2004 年 3 月.

[2] 安田浩/安原隆一監修：“ポイント図解式 コンテンツ流通教科書”，（株）アスキー，2003.7.2.

[3] 松井龍也，高嶋洋一：“電子透かしの応用：一般の利用者の協力に基づく海賊版データ摘発手法”，1998 年 暗号と情報セキュリティシンポジウム予稿集，SCIS98-10.2.C，Oct.1998.

[4] 佐竹康宏，松浦正明，松崎隆一，井上貴司，徳永裕史：“デジタルコンテンツにおける権利侵害の対処システムについて”，電子情報通信学会技術研究報告 Vol.102, No.138，DC-2002-15，2002.6.21.