

令和元年度卒業研究論文

URL の情報指向型クラシフィケーション

2020 年 2 月 7 日 (金)

指導教員 井上一成 教授

明石工業高等専門学校
電気情報工学科

報告者 E1533 西 総一郎

目次

第 1 章	序論	1
1.1	TCP/IP の課題	1
1.2	情報指向ネットワーク	2
1.3	本研究の目的	3
第 2 章	シミュレーションプログラム	5
2.1	プログラムの概要	5
第 3 章	衝突数の検証	7
3.1	ハッシュのみを用いたとき	7
3.2	URL の分類手法を利用するとき	7
3.3	ハッシュと URL の分類手法を併用したとき	7
参考文献		9

第 1 章

序論

1.1 TCP/IP の課題

1983 年から今日のインターネットと呼ばれているネットワークにおいて通信プロトコル TCP/IP がデファクトスタンダードとなった [1]. 約 20 年前のインターネットのトラフィックや利用形態は現在とは大きく異なっている. 1992 年の全世界のインターネットトラフィックは 1 日あたり約 100 GB であったが, その 10 年後の 2002 年には 1 秒あたり 100 GB に増え, 2017 年には 1 秒あたり 45,000 GB 以上に到達した. また利用形態も 2017 年においてはトラフィックの 75% をビデオコンテンツが占めている. Cisco によると全世界のインターネットトラフィックは 2022 年には 150,700 GB/秒となりその 82% をビデオコンテンツが占めると予測されている [2].

また, インターネットの使用目的も変遷している. 当初はインターネットを高性能コンピュータあるいは高性能プリンタを利用するように, 様々なリソースを遠隔から共有することが主な目的であった. 現在は情報の共有, 情報の取得といった情報のやり取りが中心となっている. それに伴って, 通信形態も変化している. 従来の TCP/IP はホスト中心の Host-to-Host の通信形態であり, IP プロトコルは位置情報であるネットワークアドレスを用いてホストアドレスを指定するというロケーション・オリエンテッド*¹な通信であった. ところが, 現在は情報をユーザに送るというインフォメーション・セントリック*²な通信形態に変わりつつある.

このように TCP/IP の通信形態と現在のインターネットに求められている通信形態との間の差が広がっている. 情報の効率的な取得のために P2P*³や CDN*⁴などの新しいプロトコルが提案された. しかし, これらはロケーション・セントリックな TCP/IP ネットワーク上のプロトコルであるので本質的な解決ではない. 本来, 情報を取得するという行為に対して, ネットワークアドレスやホストアドレスなどを意識する必要はなく, もし近くにある通信機器が当該コンテンツ (情報)*⁵を持っておりそこから情報を取得できるなら, それはより効率的であり将来の通信量増大にも対応できると考えられる. そこで, 情報を効率的に取得するために情報指向ネットワーク: Information-Centric-Network (ICN)[3] というプロトコル体系が提案された [4].

*¹ Location-oriented: 地理的指向な

*² Information-Centric: 情報指向な

*³ Peer to Peer: インターネットにおいて一般的に用いられるクライアント・サーバ型モデルでは, データを保持・提供するサーバとそれに対してデータを要求・アクセスするクライアントという 2 つの立場が固定されているのに対して, 各ピアに対して対等にデータの提供及び要求・アクセスを行う自立分散型のネットワークモデル

*⁴ Content Delivery Network: 頻繁に使われる Web サイトがある一つのノード (サーバ) だけでは耐えきれないのでいくつかのノードにデータを分散しておき, 各ユーザは分散したノードに接続して情報を取得するという方法

*⁵ 参考文献 [3] では情報 (Information) とコンテンツ (Contents) は同様の意味で用いられている. 本稿でも同様の意味で用いる.

1.2 情報指向ネットワーク

情報指向ネットワーク (ICN) においてユーザはサーバの IP アドレスではなくコンテンツ名を指定してコンテンツ取得要求を行い、そのコンテンツ要求を受け取った近隣のルータやノードが当該コンテンツを保持していた場合、それらはユーザに対してそのコンテンツを直接転送するプロトコル体系である。ICN において情報を保持している者をパブリッシャ (Publisher)、情報の取得要求を出すものをサブスクライバ (Subscriber) と呼ぶ。また ICN では各コンテンツ (情報) に対してコンテンツ名 (名前) が対応付けられている。ICN へのアプローチとして様々な研究がなされているが、現在最も多くの研究者により研究されている Named Data Networking (NDN)[5] 及びその前身である Content Centric Networking (CCN)[6] を代表的な ICN アーキテクチャとして述べる。CCN アーキテクチャはパロアルト研究所^{*6}により研究されている ICN の先駆となった本格的なアーキテクチャである。また、US Future Internet Architecture プログラム^{*7}によって資金提供された NDN プロジェクトは、CCN アーキテクチャをさらに発展させたものである。

NDN におけるコンテンツ名の命名規則は階層構造になっており、現在のインターネットで流通している識別子である Uniform-Resource-Locator (URL) に似ている。たとえば、コンテンツ名は `/aueb.gr/ai/main.html` となる。ただし、コンテンツ名は必ずしも URL とは一致せず、最初のセクション^{*8}は DNS 名または IP アドレスなどの形式である必要もない。つまり NDN では、各セクションについての具体的な規格は定義されていない。またコンテンツ取得要求において、要求されたコンテンツ名のプレフィックスの名前を持つ情報と一致すると見なされる。たとえば、`/aueb.gr/ai/main.html` は、`/aueb.gr` という名前の情報と一致する。

NDN において、サブスクライバはコンテンツを取得する際はコンテンツ取得要求である INTEREST メッセージを発行して、パブリッシャからの DATA メッセージの形式で到着するコンテンツを取得する。INTEREST/DATA メッセージは、それぞれ要求/転送される情報のコンテンツ名を持つ。Fig.1.1 に示すように、すべてのメッセージは Content Router (CR) によってホップバイホップ (hop by hop) で転送され、各 CR は 3 つのデータ構造 (Forwarding Information Base (FIB), Pending Interest Table (PIT), Content Store (CS)) がある。FIB は、INTEREST メッセージを適切なデータソースに転送するために使用する必要がある出力インターフェイスにコンテンツ名をマッピングする。PIT は、保留中の INTEREST メッセージが到着した着信インターフェイス、つまり一致する DATA メッセージが転送されてきたときに返送するインターフェイスとコンテンツ名をマッピングすることで INTEREST メッセージを追跡する。最後に、CS は CR を通過した情報のローカルキャッシュとして機能する。

INTEREST が到着すると、CR は情報名を抽出し、要求されたプレフィックスと一致する名前を持つ CS の情報オブジェクトを探します。何かが見つかった場合、すぐにデータメッセージで着信インターフェイスを介して送り返され、INTEREST は破棄されます。それ以外の場合、ルータは、この INTEREST を転送する方向を決定するために、FIB で最長のプレフィックス一致を実行します。FIB でエントリが見つかった場合、ルータは PIT に INTEREST の着信インターフェイスを記録し、FIB が示す CR に INTEREST をプッシュします。図 3 では、サブスクライバは `/aueb.gr/ai/new.htm` という名前の INTEREST を送信します (矢印 1~3)。PIT に正確な名前

^{*6} Palo Alto Research Center (PARC) : アメリカ合衆国のカリフォルニア州パロアルトにある研究開発企業

^{*7} NSF FUTURE INTERNET ARCHITECTURE PROJECT (<http://www.nets-fia.net/>)

^{*8} ”/” で区切られた部分をセクションと呼ぶ

のエントリがすでに含まれている場合、つまりこの正確な情報オブジェクトが既に要求されている場合、ルーターは着信インターフェイスをこの PIT エントリに追加し、INTEREST を破棄して、情報オブジェクトのマルチキャストツリーを効果的に形成します。

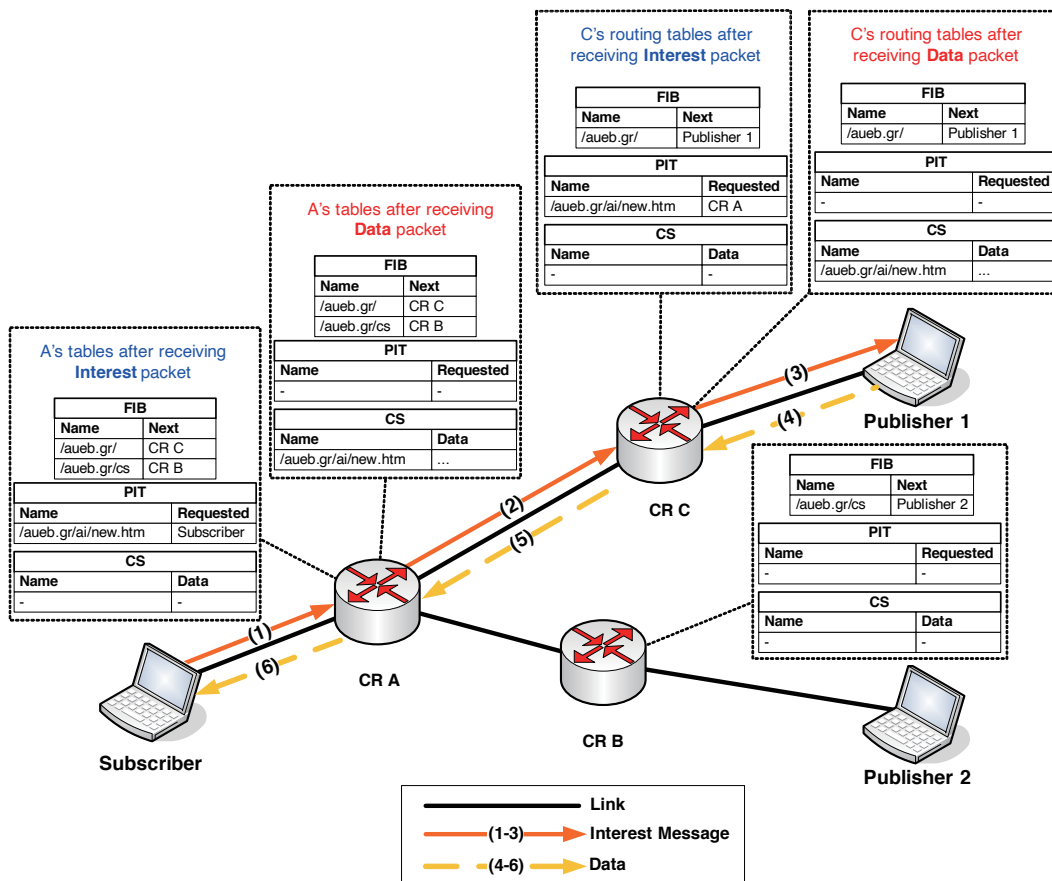


Fig1.1 The CCN/NDN architecture. CR stands for Content Router, FIB for Forwarding Information Base, PIT for Pending Interest Table, CS for Content Store (Excerpt from [7]).

1.3 本研究の目的

第 2 章

シミュレーションプログラム

2.1 プログラムの概要

第 3 章

衝突数の検証

3.1 ハッシュのみを用いたとき

3.2 URL の分類手法を利用するとき

3.3 ハッシュと URL の分類手法を併用したとき

参考文献

- [1] David D. Clark et al. Barry M. Leiner, Vinton G. Cerf. Brief history of the internet. Internet Society, 1997.
- [2] Cisco. Cisco visual networking index: Forecast and trends, 2017 - 2022. Cisco, 2019.
- [3] Van Jacobson, Diana K. Smetters, James D. Thornton, Michael F. Plass, Nicholas H. Briggs, and Rebecca L. Braynard. Networking named content. In *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '09, pp. 1–12, New York, NY, USA, 2009. Association for Computing Machinery.
- [4] 朝枝仁, 松園和久. 情報指向ネットワーク技術におけるプロトタイプ実装と評価手法. コンピュータ ソフトウェア, Vol. 33, No. 3, pp. 3_3–3_15, 2016.
- [5] NSF Named Data Networking project. [Online]. Available: <http://www.named-data.net/>.
- [6] Content Centric Networking project. [Online]. Available: <http://www.ccnx.org/>.
- [7] G. Xylomenos, C. N. Ververidis, V. A. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. V. Katsaros, and G. C. Polyzos. A survey of information-centric networking research. *IEEE Communications Surveys Tutorials*, Vol. 16, No. 2, pp. 1024–1049, Second 2014.