

## 情報の発信・共有・検索・受信に積極的に関わるセマンティックルータ

鯉 渕 道 紘<sup>†1</sup> 永 富 泰 次<sup>†2</sup> 牧 野 友 昭<sup>†2</sup>  
辻 良 繁<sup>†3</sup> 橋 岡 大 地<sup>†2</sup> 明 石 大<sup>†2</sup>  
三 野 峻 徳<sup>†2</sup> 石 田 慎 一<sup>†2</sup> 井 上 恒 一<sup>†2</sup>  
川 島 英 之<sup>†4</sup> 西 宏 章<sup>†2</sup>

インターネットにおいて流通する情報量は増加の一途を辿っており、トラフィック自体とそのコンテンツが大きな価値を持つようになってきている。そこで、我々は、パケット転送するのみならず、情報の発信・共有・検索・受信に積極的に関わるセマンティックルータを提案することで、(1) ルータが検索エンジンなどのサービスに関与し、(2) インターネットの省トラフィックを実現することを目指す。本稿ではその構想と鍵となる要素技術について述べ、先進的なインターネットバックボーンにおける実現の可能性について検討を行う。

### The Semantic Router That Actively Sends, Shares, Searches, and Receives Information

MICHIHIRO KOIBUCHI,<sup>†1</sup> YASUJI NAGATOMI,<sup>†2</sup> TOMOAKI MAKINO,<sup>†2</sup>  
YOSHISHIGE TSUJI,<sup>†3</sup> DAICHI HASHIOKA,<sup>†2</sup> DAI AAKASHI,<sup>†2</sup>  
TAKANORI MITSUNO,<sup>†2</sup> SHIN-ICHI ISHIDA,<sup>†2</sup> KOUICHI INOUE,<sup>†2</sup>  
HIDEYUKI KAWASHIMA<sup>†4</sup> and HIROAKI NISHI<sup>†2</sup>

The amount of information continues to increase in Internet, and the information of Internet traffic itself and its content are valuable for enriching services. We thus proposed a semantic router that actively sends, shares, searches, and receives information in order to (1) provide the rich information to clients and services, such as search engine, and (2) reduce the amount of the traffic in Internet. In this paper, we illustrate its concept, and the key components. We also show the feasibility of the semantic router on an existing advanced Internet backbone.

#### 1. はじめに

インターネット上に存在するあらゆる情報は、マッシュアップなど複数発信源からの情報を組み合わせる手法により、多角的な価値とコンテンツとしての意義を有するようになった。例えば Google や mixi が次々と生み出す新技術が新たなビジネスを掘り起こし広げている。このようなサービスからみた情報の融合は特に Web2.0 により大きく進歩し、今後さらにサービスをリッチにすることが望まれる。

現在、インターネット上の情報検索サービスは、巨大なストレージを持つエンドホストにより提供される。これらの検索エンジンはウェブクローラによってインターネット上の情報を収集するだけに留まっており、エンドホストから得られる情報だけに頼って検索結果の表示順位を決定している。例えば利用者があるサイトに訪れた後、どのサイトに訪れたか等の行動履歴やサイトの滞在時間に基づく検索順位情報の決定などは、順位決定に必要な情報が欠如するため困難である。従来検索サービスの中には、検索サービスのパーソナライズ化を目的とし、検索履歴を利用する検索エンジンもあるが、利用者の行動は検索キーのみで測ることができないため、行動履歴の収集には限界が生じてしまう。

一方、トラフィック情報を把握するという点では、ルータは優れた位置に存在する。これは、インターネット上の情報はすべてパケットの中継機器であるルータを介してやり取りされるためである。近年のルータは高い計算能力や交換能力の獲得により、単なる IP パケットの中継機器であるには留まらず、QoS 管理や TCP 層にお

<sup>†1</sup> 国立情報学研究所/総合研究大学院大学  
National Institute of Informatics/The Graduate University for Advanced Studies  
<sup>†2</sup> 慶應義塾大学大学院 理工学研究科  
Graduate School of Science and Technology, Keio University  
<sup>†3</sup> 東京大学大学院 情報理工学系研究科  
The Graduate School of Information Science and Technology, The University of Tokyo  
<sup>†4</sup> 筑波大学大学院システム情報工学研究科 Graduate School of Systems and Information Engineering, University of Tsukuba

よぶパケット解析が可能であり、ユーザや ISP 向けのサービス提供が進んでいる。これらの研究はルータが単なる通信基盤に留まらず、次世代のインターネットにおける情報技術とサービスの中継となりうることを示唆している。

ネットワーク上のトラフィック情報は、「ある URL にアクセスしたユーザは、他のどんな URL にアクセスするのか」、「ユーザがある URL にどのくらいの時間滞在していたのか」、「その情報がいつ、どこからネットワーク上に現れたのか」といった従来の手法では利用されなかったサービスにとって重要な情報を含む。また、これらの情報をリアルタイムに取得できる可能性がある。このような情報を、ネットワークにおける 4 大要素である、情報の検索・受信・共有・発信に関連付けることで、ネットワークサービスを圧倒的に豊かにできると考えられる。

そこで、我々は、パケットストリームをメモリデータベースに貯え SQL を拡張した SSRQL(Semantic Switch and Router Query Language) により、トラフィックに含まれるコンテンツ情報を提供するセマンティックルータを提案している<sup>1)</sup>。

セマンティックルータにより実現される代表的なアプリケーションには行動履歴解析と侵入検知がある。行動履歴解析とは、あるユーザがどのようなサイトを閲覧し、そこでどのような行動を行ったかを解析するものである。侵入検知とは、不正なアクセスの兆候を検知することである。行動履歴解析も侵入検知も、現在はサーバ上において異なる方式で実現されている。そのため同じような処理が異なるサーバ上で同時に実行されている。これは管理コスト面からも計算処理コスト面からも無駄である。これらはセマンティックルータにより実行することにより両問題は解決される。さらに、インターネット上に頻繁に転送される情報については、セマンティックルータのストレージにキャッシングすることでトラフィック量を削減することができる。

以下、2 章で本研究の関連研究を紹介し、3 章においてセマンティックルータのアーキテクチャを述べる。4 章においてセマンティックルータが関与するサービスを示し、5 章においてセマンティックルータを我が国の学術情報ネットワークである SINET3 に適用するための検討を行う。最後に 6 章でまとめを述べる。

## 2. 関連研究

サービスやコンテンツと親和性の高いネットワークの構築を目指し、新しいルータアーキテクチャが提案されている<sup>2),3)</sup>。これらは、既存の一般的なルータがパケットヘッダを解析することによってルーティングを行うのに対し、パケットペイロードに含まれる情報をもとにルーティングを行うことを志向している。このようなルータ

とパケットペイロードの新しい相互作用を創出する研究は、コンテンツベースルーティングのアーキテクチャ研究において議論されている。また、構造化された XML パケットに特化してコンテンツベースルーティングを行う XML ルータは商用化され、利用が始まっている。

一方で、我々は、トラフィックデータを解析あるいは必要に応じて蓄積し、それをコンテンツとしてサービスに提供するサービスとの連携を行う。そして、そのサービスとのインタラクションを通じてルーティングに活用できるコンテンツ情報をコンテンツベースルータのルーティングテーブルとして利用する。

## 3. セマンティックルータのアーキテクチャ

既存のインターネットルータはデータ転送とプロトコル変換という機能の高度化を追求したハードウェアである。それに対して、我々が開発しているセマンティックルータは、単にデータ転送とプロトコル変換を行うハードウェアではない。セマンティックルータは、トラフィックデータ中のペイロードを含めたパケットを処理し、web サービス、クライアント、周辺ルータが利用可能な形で情報を提供する。

従来のルータはプロトコルを処理するのみであるのに対し、セマンティックルータはパケットペイロード中のコンテンツデータを参照し、より豊富な情報でルーティングテーブルを定義してインターネットトラフィックの有用性と管理機能を向上させることができる。さらに、行動履歴解析や侵入検知を含めた幅広い処理をユーザに提供する。インタフェースは記述能力の高い宣言的言語である。これにより、従来は不可能であった、細粒度かつ広範囲の行動履歴解析および、侵入等の攻撃に対するルータでの一括防御を支援する。

セマンティックルータの構成を図 1 に、その中のパケット処理エンジンを図 2 に示す<sup>4)</sup>。ワイヤレト処理が必要となるトラフィックパケットのパターンマッチング処理やデータベースへの書き込み処理は、それぞれパケット処理エンジン、および DB Insertion Engine といった専用ハードウェアにより高速処理される。正規表現プロセッサはパケット処理エンジンの中に含まれており従来のネットワークプロセッサに内包されるプロセッシングユニットのコプロセッサとして実装される。この正規表現プロセッサは、専用のクエリ言語である SSRQL によるインサージョンの中でも複雑な正規表現マッチングを並列して実行する。一方、ワイヤレトを必要としないデータベースからのデータ取得は SSRQL 処理エンジンによってソフトウェアで処理される。

図 2 に示したパケット処理エンジンでは、まず、MAC/IP Extractor において、ルータ外部から入ってきたパケットから IP アドレスと MAC アドレスを抽出する。TCP Stream Reconstructor では、TCP パケッ

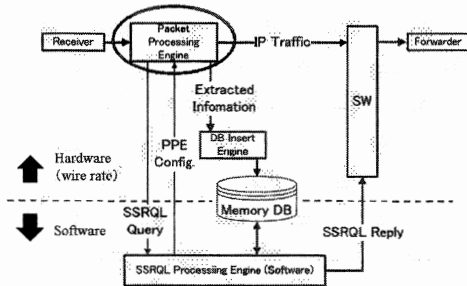


図1 セマンティックルータの構成

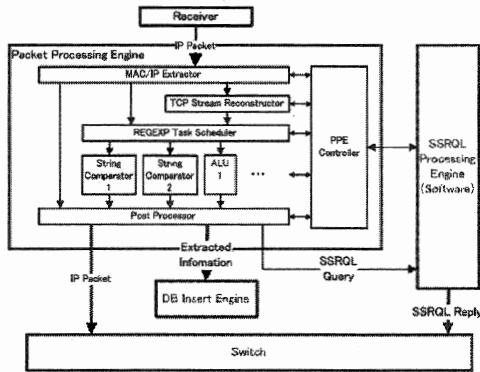


図2 パケット処理エンジン

ともしくは5タプルとSource Port IDに従ってストリームを再構成する。次に、通常のパケット処理を行うためにパケット処理エンジンと呼ばれるプロセッシングユニットアレイにパケットが割り当てられる。同時に正規表現処理専用コプロセッサが起動される。REGEXP Task Schedulerは複数存在する正規表現プロセッサにストリームを割り当てる。正規表現プロセッサは専用のマイクロコードにより正規表現のマッチング処理を行う。正規表現パターンにマッチし、抽出対象となった情報はPost Processorを介して、DB Insertion Engineによりデータベースに蓄えられる。

現在、セマンティックルータは設計中であるが、今後、(1)ワイヤレート (Gbps~10Gbps/リンク) で上記処理ができることを確認し、(2)実インターネットトラフィックトレースを用いて必要となるメモリ量などの定量的な評価を行い、(3)セマンティックルータにより新たに提供することが可能となった情報の解析、ならびに機能を加えることで省トラフィック量についての解析を行う予定である。

#### 4. セマンティックルータが関与するサービス

SINET<sup>3</sup>などの既存のインターネット、そこに上位

レイヤのサービスを付加した P2P オーバレイネットワーク、セマンティックルータにより構築されたインターネットの比較を表1に示す。

	既存の NW	P2P	Semantic
NW/コンテンツ情報	×	×	○
省トラフィック	×	○	○
セキュリティ	×	×	○
情報の耐災害性	×	○	○

#### 4.1 リッチな NW 情報

1章、3章の議論より、セマンティックルータは、詳細な行動履歴解析などの、よりリッチなネットワーク・コンテンツ情報をクライアント、アプリケーションに提供することができる。一方、既存のインターネット、P2P オーバレイネットワークでは、クライアント、アプリケーションから物理的なネットワーク構成を知ることは通常できない。

#### 4.2 省トラフィック

セマンティックルータを用いたインターネットでは、頻繁にアクセスされるデータのレプリカを距離の近いルータに格納することでトラフィック量を削減できる可能性がある。同様のことは P2P オーバレイネットワークでもいえるが、セマンティックルータと組み合わせることで、P2P トラフィックに介入し、下位層を無視した上位でのルーティングではなく、SR が提供する情報を利用した下位層を正確に反映した P2P が可能となる。つまり、ユーザは、余っているディスク領域を使って自由にキャッシュを持ち、その管理をセマンティックルータが行う制御が可能である。

#### 4.3 セキュリティ

ネットワーク自らが不正アクセスの侵入検知に関わるため、他のネットワークと比べ、セキュリティを高める技術の開発が容易となる。なお、プライバシー情報についての管理もセマンティックルータは行うが、本稿では詳細は省略する<sup>1)</sup>。

#### 4.4 情報の耐災害性

セマンティックルータはホストのもつ情報と一貫性を保証することはできないが、データのレプリカをルータが格納することができるため、ディザスタグリッドと同様に情報の耐災害性を持つことができる。

### 5. 先進的なインターネットへの適用

現在、学術ネットワークなどの先進的なインターネットでは、単なる IPv4 や IPv6 native の通信のみならず、ネットワークのサポートによる VPN を用いたセキュアな通信等高度な機能が付随している。本章では、このような先進的なインターネットにおいてセマンティックルータを適用した場合の検討を行う。

例えば、約 700 機関のユーザを抱える我が国の学術情



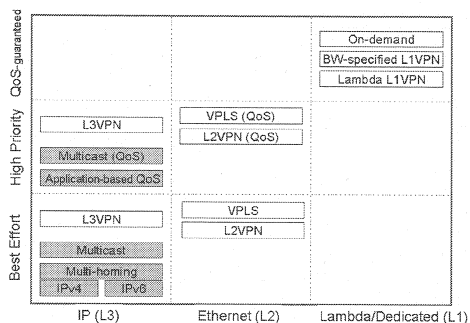


図3 SINET3のマルチレイヤサービス

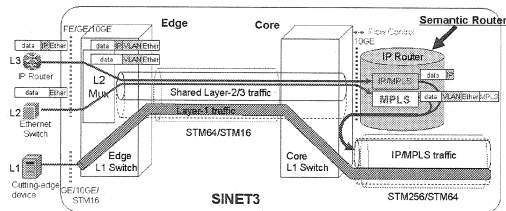


図4 マルチレイヤサービスの収納

報ネットワーク SINET3 では、ユーザ、アプリケーションの要求に合わせて、図3のように様々なマルチレイヤサービスを提供している<sup>5)</sup>。

この中で、セマンティックルータが情報の共有・検索に関与することができる通信は、図3の斜線の通信サービス(IP(L3)通信)のみである。ただし、SINET3においてセキュリティを高めるために閉域を用いる通信(VPN)やL1 オンデマンドのように通信帯域を保証する通信は、通信相手が事前に決まっておき、利用内容についてネットワーク運営者に申請している場合がほとんどである。そのため、セマンティックルータの収集の対象となる情報は、そもそもほとんどないと考えられる。

次にセマンティックルータの設置について考える。現在、SINET3では図4のようにVLAN技術を用いてレイヤ2/レイヤ3層の通信を多重化している。レイヤ1の通信は各大学などのノード、データセンターにあるL1スイッチを用いて転送しているためバックボーンルータは関与していない。そのため、このバックボーンルータ(現在、Juniper社T640)をセマンティックルータに置き替えることで、現状のSINET3のアーキテクチャを維持しつつ、新たにセマンティックルータのサービス展開が実現可能である。よって、SINET3の場合、全国12箇所にあるデータセンターに各1台、計12台のセマンティックルータが必要ということになる。

さらに、セマンティックルータにより新たに提供できるリッチ情報(行動履歴解析等)は、SINET3の場合、運営主体の国立情報学研究所が提供している上位アプリケー

ションにおいて利用可能である。つまり、セマンティックルータによりネットワークインフラストラクチャとそのアプリケーションの1つである学術コンテンツ、検索を一体化した密な学術基盤(サイバーサイエンスインフラストラクチャ:CSI)のパッケージングがより容易かつリッチになるといえる。

## 6. まとめ

インターネットにおいて流通する情報量は増加の一途を辿っており、トラフィック自体とそのコンテンツが大きな価値を持つようになってきている。そこで、我々は、データ転送するのみならず、情報の発信・共有・検索・受信に積極的に関わるセマンティックルータを提案することで、(1)ルータが検索エンジンなどのサービスに関与し、(2)インターネットの省トラフィックを実現することを目指している。本稿ではその構想と鍵となる要素技術について述べ、先進的なインターネットバックボーンにおける実現の可能性についても検討を行った。

今後は、この構想に基づいて、セマンティックルータの設計を完了させ、(i)インターネットバックボーンの回線速度(Gbps~10Gbps/リンク)で動作できることを確認し、(ii)実インターネットトラフィックトレースを用いて必要となるメモリ量などの定量的な評価を行い、(iii)セマンティックルータにより新たに提供することが可能となった情報の解析、ならびに機能を加えることで省トラフィック量についての解析を行う予定である。

## 謝 辞

本研究の一部は、独立行政法人・情報通信研究機構(NICT)の委託研究「新世代ネットワーク構成に関する設計・評価手法の研究開発」の支援による。また、本研究はSINET3の運用、設計とは無関係である。

## 参 考 文 献

- 1) Inoue, K., Akashi, D., Koibuchi, M., Kawashima, H. and Nishi, H.: Semantic router using data stream to enrich services, *In Proc. 3rd Intl. Conf. on Future Internet (CFI)*, pp. 20-23 (2008).
- 2) Yang, C.-S. and Luo, M.-Y.: Efficient support for content based routing in web server clusters, *In Proc. USENIX Symposium on Internet Technologies and Systems (USITS)* (1999).
- 3) Carzaniga, A., Rutherford, M. J. and Wold., A. L.: A routing scheme for content-based networking, *In Proc. IEEE INFOCOM* (2004).
- 4) 永富泰次, 石田慎一, 三野峻徳, 川島英之, 鯉淵道紘, 西宏章: リッチなユーザサービスを提供するセマンティックルータにおける正規表現プロセッサの提案, 電子情報通信学会, ネットワークシステム研究会 (NS) (2008).
- 5) Urushidani, S., Abe, S., Fukuda, K., Matsukata, J., Ji, Y., Koibuchi, M. and Yamada, S.: Architectural Design of Next-generation Science Information Network, *IEICE Transaction*, No. 5, pp. 1061-1070 (2007).