

ビッグデータ時代におけるインターネットのあり方の一考察 ～エコという視点から

菅原彬 鈴木淳史 田中一哉 周鵬

(アドバイザー教員 片岸一起)

1 はじめに

近年におけるインターネットの技術の発展は、人類史上稀に見る進化であり、もはやインターネットが無くては成り立たない社会になっている。2012年におけるインターネットのユーザ数は22億3千万人とされており、2007年と比べ約2倍に増加している[1]。

現在もインターネットは成長し続けており、ユーザ数も増加し続けている。これに加えてスマートフォンなどの普及に伴いインターネットに接続される機器も増加している。ユーザの増加は、トラフィックを増加させ、さらにはそれを扱う情報機器の消費電力も増加させる。そこで本論文ではインターネットのエコとして、トラフィックと情報機器消費電力の削減を扱い、ビッグデータ時代におけるインターネットのあり方について考察を行う。

本論文は、まず第2章で現在までのインターネットの変遷について述べる。次に第3章では、エコの視点から考えられる現行インターネットの問題点について挙げる。また、その問題を引き起こしている一因と考えられるもののケーススタディを行う。第4章では、これからのネットワークの形として考えられるものとして、現行インターネット上で構築できる Peer to Peer(P2P)ネットワークと、現行インターネットに代わる新世代ネットワークについて述べる。また、ケーススタディにおけるエコに向けた考察を行う。最後に第5章でまとめとして、ビッグデータ時代においてインターネットはどうあるべきか考察を行う。

2 インターネットの変遷

本章では、インターネットがビッグデータ時代である現在まで、どのように変遷したかを概説する。

2.1 インターネットの成り立ち

現在のインターネットの原型となったのは、1969年にアメリカ国防総省により構築された研究用ネットワーク“ARPAnet”である。ARPAnetでは、初め

てパケット通信と呼ばれる通信手段を実現した。その後、Transmission Control Protocol/Internet Protocol (TCP/IP)システムの確立、World Wide Web (WWW)の実現など、今日のネットワークの基幹となる技術が次々と生まれた[2]。

2.2 インターネットの普及・成長

1990年代へ入ると、急速に情報機器が普及した。初めは、業務用として企業にパーソナルコンピュータ(PC)の導入が進み、次第に一般にも広まっていった。日本では、1994年に初の個人向けインターネット接続サービスが開始された。日本において、インターネットという言葉が一般的になったのは、翌年1995年に発生した、阪神淡路大震災でインターネットの有用性が広く知られたことで、その年の流行語大賞に選ばれたという背景がある。

この頃に、様々なインターネットサービスが誕生した。1996年には、Hotmail(電子メールサービス)、インターネットラジオ、テレビのデジタル放送、電子書籍店パピレス等、様々なコンテンツ(テキスト、音声、画像、動画等)がデジタル化された。その一方では、既存のアナログな通信網が廃止になるケースもあった。記憶に新しいものでは、テレビ放送が2011年に地上デジタル放送へ完全移行し、アナログ放送の停波が行われたことが挙げられる。

豊富なインターネットコンテンツが誕生した背景には、デジタル回線網の世界的な整備がある。日本においては、1988年に国内初の専用デジタル回線であるISDN回線が開通した。さらに、1999年のADSL回線開通と、2001年の光回線の開通を経て爆発的に普及し、2012年には回線契約数が3500万件に及んでいる。デジタル回線網の発展と情報機器の普及は、インターネットサービスの発展を後押しし、既存のあらゆるコンテンツをデジタル化へと促した。

このようにインターネットは、研究用ネットワークとして始まり、商用化に伴った一般事業の参入により、当初考えられていたよりも遥かに巨大なデータ通信網の形態を確立し、現在のインターネットへととなった。

2.3 情報爆発とビッグデータ

インターネットは発展と共に、既存の通信網や情報をデジタル化して集約し続けている（図 1）。これまでは、データ通信網、放送網（テレビ放送・ラジオ放送）、電話網は、それぞれ専用機器と回線を持ち独立した状態で存在していた。また、手紙や書籍、動画なども従来の流通形態のみで無くインターネットを利用した、デジタルコンテンツとして流通するようになった。あらゆる形態の情報を、ひとつの回線上で扱えるインターネットは、その快適さから次第にデータ量が大きくなっていき、その結果“情報爆発”へと至った[3]。

情報爆発は、現在の手法ではデータ量が大きすぎて、扱いきれない程のデータを言い表したネガティブな言葉である。近年、この膨大なデータを有効利用することで新たな価値を見出そうという流れが起きている。これまでと一転して、ポジティブな意味を持つようになった膨大な情報は“ビッグデータ”と呼ばれるようになった。

ビッグデータ時代では、大量の情報端末やセンサーなどから、膨大な情報がリアルタイムに生成され続けている。企業が収集し蓄積するデータ量も急増し、SNSの普及により非定型なデータがインターネットには溢れている。昨今のIT情勢において、ビッグデータの活用が、社会に与えているインパクトは非常に大きい。

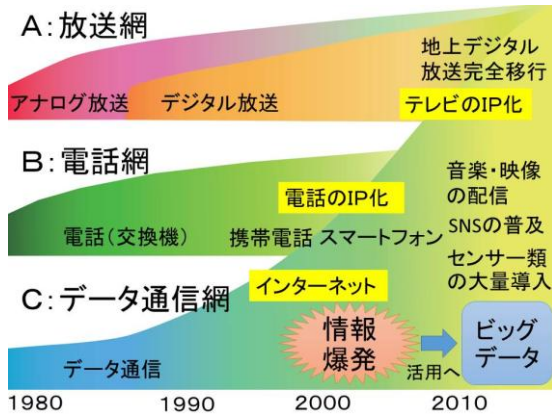


図 1. 現代までのインターネットの変遷

3 現行インターネットの問題

東日本大震災以降、日本における電力需要は逼迫した状況下であり、社会の節電意識は一層高まっている。本章では、その中で現行インターネットをエコという視点から考えたときの問題点を 3.1 節、3.2 節で挙げる。また、その問題の一因と考えられるも

のとして 3.3 節でスパムメール、重複トラフィックを挙げる。

3.1 ビッグデータ時代におけるトラフィック過多

ビッグデータ時代におけるインターネットでは、ユーザ数の増加やスマートフォンの普及、センサー類の大量導入など様々な要因により、トラフィック過多の問題が起きている。今後、日本国内のインターネットのトラフィックは、121Tbps まで増加すると予測されており、これは 1 秒間に DVD を 1 万 4 千枚消費することに相当する（図 2）。このトラフィック過多は、情報機器消費電力の増加などの問題を引き起こす一因となっている。

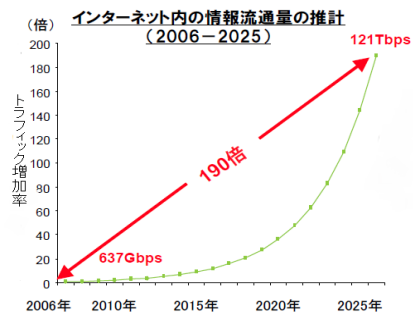


図 2. 日本国内トラフィック予測[4]

3.2 トラフィックと情報機器消費電力

ルーター等の情報機器は自身を流れるトラフィックに応じて稼働する。そのため、トラフィックと消費電力は一般に比例する。2025年時点で情報機器消費電力は、2006年時の約 5.2 倍まで増加し、国内総発電量の約 20% を占めると予測されている（図 3）。

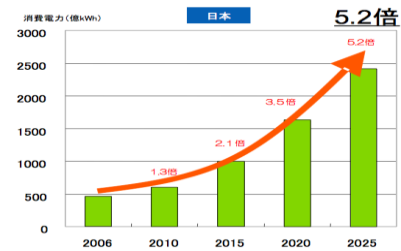


図 3. 日本の情報機器消費電力予測[4]

この日本国内情報機器消費電力予測に対し、情報機器の内訳も予測されている。2006年時点では、サーバは情報機器消費電力の内、約 3 割を占めている（図 4）。これは、現在のインターネットにおいてクライアントサーバ(C/S)ネットワークが一般的であり、サーバに特に負荷がかかるためであると考えられる。

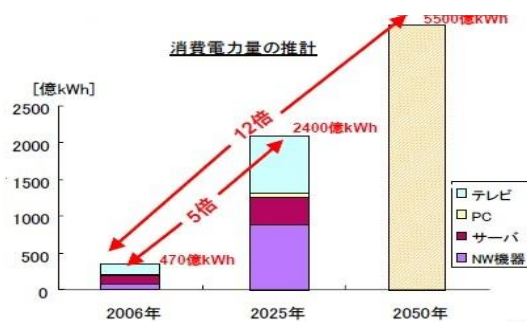


図4. 情報機器消費電力の内訳[4]

3.3 ケーススタディ

本節では、3.1節、3.2節で述べた現行インターネットの問題を引き起こしている一因として考えられるスパムメール、重複トラフィックを挙げる。

3.3.1 スパムメール

スパムメールの歴史は古く、手紙の役割が電子メールとしてインターネットに代替された頃からあったが、近年ではメールユーザーの増加に伴いさらに増えている。世界におけるメールの利用者数は、2012年時の推計で22億人とされており、世界の約3人に1人がメールユーザーである[5],[6]。また、サイバー攻撃の起因となるようなスパムメールが広がりを見せていることから、スパムメールは電子メールにおいて優先的に解決すべき問題である。

3.3.1.1 スパムメールとは

スパムメールとは受信者の意向を無視して、無差別かつ大量に一括送信される、電子メールを主としたメッセージのことである[7],[8],[9]。電子メール利用者が、事前に許可していない広告メールや出会い系サイトの勧誘などが挙げられる。

3.3.1.2 スパムメールの消費電力

コンピュータセキュリティ大手の McAfee, Inc. とコンサルティング業務を行う ICF International は、2008年にスパムメールに関する世界的な調査[10]を行った。調査手法は、大手プロバイダや企業の協力の下で受信者のいないアドレスを約300万個用意し、そのおとりアカウントに受信するメールを独自の手法を用いて集計するというものである。当然、受信者のいないアドレスに届くメールは、通常のメールではないためスパムメールとなる。

その結果、スパムメールの1日あたり配信数は約1750億件と分かった。また、調査の年に世界中で送信されたスパムメールは合計で62兆件であり、この数は電子メールの総数77兆件[8]に対し、約9割もの電子メールがスパムメールだったことを表している。

このことは、メール使用者にとって通常のメールよりも、スパムメールの方がはるかに多く届くといった事態になり、大きな社会問題となっている。

スパムメールに関わる消費電力量を調査するため、McAfee, Inc. と ICF International では、スパムを受信者が見るまでの各段階でのエネルギー消費量を計測した(図5)。その結果、全消費電力中の9割をスパムメールの表示とフィルタリング、誤検知が占めていることが分かった。

年間のスパムメールの総数62兆件と上記の調査により分かったメール1通辺りの消費電力2.7mWを掛け合わせると、スパムメールに関わる年間のエネルギー消費量は、世界全体で165TWhに及ぶ事が分かった。その内、135TWhに関しては、既に導入されているスパムフィルタによって節約されていること報告されている[10]。よって、実際に消費されているスパムメールに関わる年間のエネルギー消費量は、世界全体で33TWhになると算出される。これは、米国240万世帯が1年間に消費する電力量に相当する。

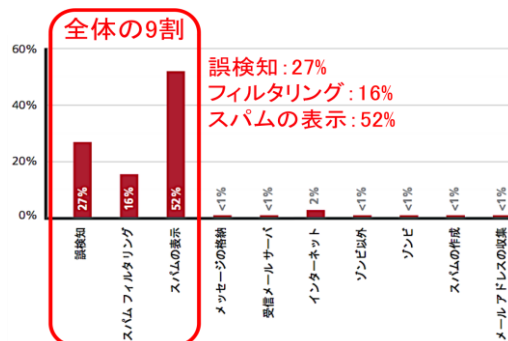


図5. スパムメールの消費電力内訳

3.3.2 重複トラフィック

近年、YouTube等に代表される動画共有サービスの利用が拡大し、動画トラフィックが増加している。これに関し、動画トラフィックは2012年時点で全トラフィックの約6割を占めている[11]。しかし、動画トラフィック中には、同一な動画に関するトラフィックが数多く存在しており、これの削減がトラフィック過多の軽減、及び情報機器消費電力の削減に繋がると考えられる。

3.3.2.1 重複トラフィックとは

重複トラフィックとは、同一な内容の動画を要求した際に発生するトラフィックのことである[12]。動画共有サービスの人気ランキング上位の動画に関し、重複トラフィックは特に発生しやすいと分かる。

3.3.2.2 重複トラフィックの消費電力

2017年時点でインターネットの全トラフィック

の約7割が動画トラフィックであると予測されている[11]。これに関し、動画トラフィック中にどの程度重複トラフィックが含まれているのか解析が行われており、あるキャンパスネットワークでは、YouTubeの動画トラフィックの約3割が重複トラフィックであることが分かっている[12]。ここで動画トラフィックの約3割が重複トラフィックであると仮定する。これよりインターネットの全トラフィックの約2割が重複トラフィックとなる。トラフィックと情報機器消費電力は比例関係にあり、仮にルーターの消費電力のうち約2割が重複トラフィックによるものとする、2017年時点でその消費電力は約200億KWh[13]と予測され、玄海原子力発電所の発電量と同等な電力を消費すると試算される。

4 これからのインターネット

これからのインターネットでは、スパムメールや重複トラフィックといった余分と考えられるデータを流さない、ということがトラフィック過多の軽減や機器の消費電力削減、といった問題の軽減に繋がる。このエコに繋がると考えられるネットワークの仕組みとして、P2Pネットワーク、新世代ネットワークを挙げる。また、スパムメール、重複トラフィックをどのように軽減するのか考察を行う。

4.1 P2P ネットワークの活用

4.1.1 P2P ネットワークとは

P2Pネットワークとは、クライアントサーバ(C/S)ネットワークと異なり、ネットワーク上の他の機器に対して、クライアントとしてもサーバとしても動作する機器(ピア)の集合によって形成される通信ネットワークである。P2Pネットワークを形成する技術群を総称して「P2Pネットワーク技術」と呼ぶ[14]。P2Pネットワークはサーバを持たないPureP2Pが基本だが、最近では小規模なサーバを持つHybridP2Pの研究も進められている[15]。P2Pネットワークでは、ファイル共有の際にインターネットウイルスの拡散が容易であったり、著作権を侵害してしまうなど問題は多く存在する。

4.1.2 P2P ネットワークによるエコ

P2PネットワークはC/Sネットワークと比べ、サーバを用いないため、サーバを稼働させるための消費電力を必要としない。しかし、サーバを用いないことでピアの負担が増加するため、結局C/Sネットワークと消費電力は変わらないのではないかと、という議論もある。いまだC/SネットワークとP2Pネッ

トワークの消費電力に関して比較は行われていないが、P2Pネットワークについてトラフィック削減や効率化を実現するための技術がいくつか提案されている。以下では、この技術について概説する。

4.1.2.1 Proactive network Provider Participation for P2P(P4P)

P4P技術はP2Pネットワークのピア選択方法を改善したものである。P2Pネットワークでは、あるコンテンツを要求する際に、そのコンテンツを持つピアを選択する必要がある。その際に、自身と同じインターネットサービスプロバイダ(ISP)内に所属するピアではなく、外部のISPに所属するピアを選択する場合がある。P4PではISPからピアの情報を得てなるべく自身に近いピアを選択するようにすることで、より効率の良いピアの選択を可能にしている[16]。このP4P技術はトラフィック削減に関し、その有効性を示す実験が行われている[17]。

4.1.2.2 ISP間トラフィック削減技術

P4P技術では、コンテンツを要求する際になるべく自身の近くのピアを選択することで、トラフィックを削減したが、通信速度は必ずしも速くなるわけではない。そこでユーザの利便性を考慮し、通信速度とトラフィックの両方を考慮に入れた指標に基づきピアを選択する手法が提案されている[18]。この手法では、ダウンロード時間の増加を極力抑え、トラフィックを削減することも可能にしている。

4.2 新世代ネットワークの活用

4.2.1 新世代ネットワークとは

第3章で述べたとおり、トラフィック過多とこれに伴う消費電力の増加はインフラとしてのインターネットの信頼性に深刻な影響を及ぼしている。そこで、このような状況に対応するため、現行のインターネットに代わるネットワークを作り出そうという動きが世界各国で始まっている。これが「新世代ネットワーク」構築構想である[19]。

4.2.2 新世代ネットワークによるエコ

新世代ネットワークを実現する技術の中で、エコと関連し注目すべきものとしては、「オール光化技術」「光パケット・光パス統合技術」「ネットワーク仮想化技術」が挙げられる。以下では、これら各技術について概説する。

4.2.2.1 オール光化、光パケット・光パス統合技術

現在のインターネット上で使われているルータは、光回線に繋がっている場合でも、光信号をいったん電気信号に変える処理を行っている。これを光のまま処理しようとするのが「オール光化技術」である

[20]。これにより光信号を電気信号に変換し、さらにまた光信号に戻すというプロセスを省略でき、変換による消費電力を削減することができる。

一方、「光パケット・光パス統合技術」とは、パケットを光信号化し、そのままルーティングを行う「光パケット交換」と、ノード間で事前にパスを設定し、中継ノードにおいて予め設定された出力インターフェースに光信号のまま交換する「光パス交換」とを同一ネットワーク上に併存させる技術である。その上で、両者をトラフィックの変化などに応じて動的に使い分けることで、通信回線の効率的利用が図られ、ネットワーク全体の省電力化に繋がる[21],[22]。

4.2.2.2 ネットワーク仮想化技術

仮想化とは、元来、「システムやネットワークを構成する各種の物理資源を、あたかも論理的に異なる資源であるかのように認識させる技術」である。この技術は、サーバからストレージへとその適用範囲を拡げてきたが、近年、これがネットワークにまで及びつつある。その具体例の1つが、新世代ネットワークにおけるネットワーク仮想化技術である。

新世代ネットワークにおけるネットワーク仮想化は「リンク資源だけでなく、ノード上の計算資源、ストレージ資源をも含むインフラストラクチャ全体の資源の仮想化」と定義される[23]。ここではネットワーク上の資源の集合を「スライス」と呼ぶ(図6)。また、このスライスを実現するためのネットワーク機器を仮想化ノードと呼ぶ。ユーザはこのスライス上に自分の目的に即したネットワークを自由にプログラミングすることができる[24]。

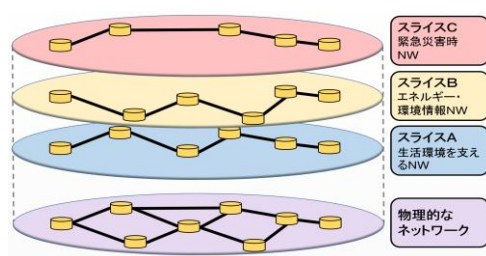


図6. 新世代ネットワークのスライス

すなわち、新世代ネットワークでは、ユーザはネットワーク上に存在する様々な資源を統合管理し、提供しようとするサービスに応じて、通信方式、速度、品質、機能等を設定して独自のネットワークを構築することができる。このようにして作り出された仮想ネットワークは物理ネットワーク上に多数併存し、相互に干渉し合うこともない。このようなネットワーク構築における自由度の大幅な向上により、ネットワーク資源の効率的利用が促され、よりいっ

そうの省電力化が実現できるものと期待されている。

4.3 ケーススタディにおけるエコに向けた考察

本節では、第3章で扱ったスパムメールと重複トラフィックをどのように軽減していくか考察を行う。

4.3.1 スパムメール軽減に向けた考察

スパムメールへの対処は、フィルタリングを導入することが効果的である。3.3.1.2節にて示したように、既に導入されているフィルタリングによって、スパムメールに関する電力は年間で135TWh節約されている。しかし、スパムフィルタによる誤検知を解決するために、ユーザの操作(スパムの閲覧、スパムの隔離、誤検知された通常のメールの復帰)によってスパムに関する年間の消費電力である33TWh中、27%もの電力を使用している。このことから、フィルタリングの精度の向上が求められており、誤検知の減少は消費電力の削減に繋がる。また、ユーザの操作の減少は、スパムの表示にかかる電力消費量52%の削減にも効果的である。

そこで、注目したのがビッグデータの解析手法である。ビッグデータの解析手法は、非定形型の膨大なデータを解析することに優れている。メールは全世界で一日に600億通以上送信されており、膨大な非定形型のデータである。

実際に、企業向け電子メールセキュリティを業務とするProofpointでは、ビッグデータの手法を用いたフィルタリング技術の開発を行っている。日常のメールを分析して傾向を把握することによって、それ以外のメール(スパム)を取り除く手法を研究中である[25]。これまでは、悪質なメールのみをモデル化してきたが、この手法では通常のメールも分析してモデル化している。5年前程までは、メールのビッグデータを扱う技術が十分ではなく、非現実的だったがビッグデータに関する技術を使うことによって比較的容易に実現出来るようになった。実際にProofpointでは、既に解析を始めており、一日に数十億通のメールを解析している。この処理を個人ベースでも行えるようにすることによって、より高度なフィルタリングが可能になると期待されている。

4.3.2 重複トラフィック軽減に向けた考察

重複トラフィックを軽減するためには、データをユーザ付近からダウンロードすることが必要である。これに関し、パケットキャッシュという技術が提案されている(図7)。これは新世代ネットワーク上で実現できる技術で、一度通過したデータを仮想化ノードに一時保存(キャッシュ)させることで、二回目以降の要求の際にデータを端末に近い仮想化ノ

ドから取得することが可能となる[12]。これにより、従来のトラフィックの3分の1から8分の1に抑えられると試算されている[26]。

パケットキャッシュでは、この仮想化ノードにどの程度ストレージ容量を持たせるかが問題になる。これに対し、リクエストの増減に応じ、最適なストレージ容量の算出が可能なのではないか、と考えている。リクエストの増減に関し、既にユーザの生活行動とトラフィックの関係が解析されており、例えばテレビのゴールデンタイム時に動画トラフィックが減少するといった知見が得られている[27]。

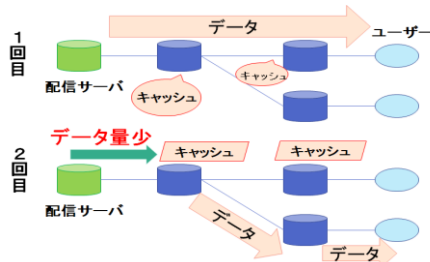


図7. パケットキャッシュ

最適なストレージ容量を算出し、ストレージ容量を動的に変化させ、使用しないストレージ容量は他のスライスで利用、もしくは休止させることでより効率の良いネットワーク運用に繋がるだろう。

5 まとめ

インターネットは想定外の成長を遂げ、ビッグデータ時代である現代においてはインターネットに関連する情報機器が電力を大量に消費しており、エコなネットワークが必要とされる。エコを実現するネットワークとして、現行インターネット上で構築するP2Pネットワーク、現行インターネットに代わるものとして新世代ネットワークを取り上げた。また、インターネットにおける余分なものとしてスパムメール・重複トラフィックを取り上げた。

エコという視点からこれからのインターネットのあり方を考えると、より効率の良いネットワークとして必要最小限のデータのみを流す、つまり余分なデータを流さないことが必要である。余分なデータを流さないことで、トラフィックが減少し情報機器の消費電力削減に繋がる。余分なデータを流さないためには、データが流れる経路を短くする、ユーザの近くからデータを収集するなど、様々な処理が考えられる。本論文で扱ったスパムフィルタリングでは、データをユーザまでなるべく届けないようにし、パケットキャッシュでは、ユーザの近くからデータ

を収集することになる。

今後は、本論文で取り上げたスパムメール、重複トラフィック以外にもケーススタディを行い、インターネット中の余分なものはなにか明確にする必要がある。そのうえでP2Pネットワークや新世代ネットワーク上での特有な処理を行えば、よりエコなネットワークへと繋がるのではないかと考えられる。

参考文献

- [1]Internetworldstats.com, (<http://www.internetworldstats.com/surfing-jp.htm>)
- [2]林井純, 「インターネット」, 岩波新書, 1995.
- [3]林井純, 「インターネット新世代」, 岩波新書, 2010.
- [4]経済産業省商務情報政策局, 「グリーンITイニシアティブ」, 経済産業省グリーンIT推進協議会資産, 2012.
- [5]Internetworldstats.com: (<http://www.internetworldstats.com/surfing-jp.htm>) (2013/09/20).
- [6]Internet 2012 in numbers: (<http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers/>) (2013/09/20).
- [7]RoyalPingdom, (<http://royal.pingdom.com/>) (2013/09/20).
- [8]Wikipedia, 「スパム(メール)」, ([http://ja.wikipedia.org/wiki/E3%82%B9E3%83%91%E3%83%A0_\(%E3%83%A1%E3%83%BC%E3%83%AB\)](http://ja.wikipedia.org/wiki/E3%82%B9E3%83%91%E3%83%A0_(%E3%83%A1%E3%83%BC%E3%83%AB))) (2013/09/20).
- [9]Microsoft, 「“スパムメール”とは何のことですか?」 (<http://support.microsoft.com/kb/882871/ja>) (2013/09/20).
- [10]McAfee, “スパムメールと二酸化炭素排出量 (McAfee, Inc. ICF)”, 2009.
- [11]Cisco, 「ゼタバイト時代とトレンド分析」, 2013.
- [12]安藤翔伍, 中尾章宏, 「ネットワークエッジにおける重複トラフィック削減のための In-Network キャッシュの検討」, 電子情報通信学会, 2012.
- [13]並木周, 挾間壽文, 「光パス・ネットワーク構想」, (http://www.aist.go.jp/aist_j/aistinfo/aist_today/vol08_07/special/p04.html), (2013/11/15).
- [14]総務省, 「P2Pネットワークの現状ネットワークの中立性に関する懇談会P2P-WG」, 2007.
- [15]総務省, 「P2Pネットワークの現状と将来」, 2006.
- [16]野田健, 「資源探索範囲を考慮したP4Pネットワークの性能評価」, 平成20年度特別研究報告, 2009.
- [17]村松謙, 「P4PによるP2Pの効率化」, (http://www.net.c.dendai.ac.jp/~murasatsu/p2p_p4p.html), (2013/11/09).
- [18]今泉友輔, 今泉貴史, 「P2PダウンロードにおけるISP間トラフィック削減手法の提案」, 学術情報処理研究 No.17, 2013.
- [19]新世代ネットワーク研究開発連絡本部, 「新世代ネットワーク技術戦略中間報告書」, 独立行政法人 情報通信研究機構, 2009.
- [20]@IT, 「オール光は省エネルギーも実現」 (<http://www.atmarkit.co.jp/news/200811/07/nict.html>), (2008/11).
- [21]建設電気技術協会, 「フォトニックネットワーク」, Kensei Denki Gijyutsu, vol.151, 2005.
- [22]和田尚也, 原井洋明, 「世界初, 光パケット・光パス統合ネットワーク基盤技術を実証」 (<http://www.nict.go.jp/press/2010/06/16-1.htm>), NICT, 2010/6/16.
- [23]中尾章宏, 「新世代ネットワーク構想におけるネットワーク仮想化」, 電子情報通信学会誌 Vol.94, No.5, 2011.
- [24]富永昌彦, 「2020年頃の将来社会の基盤となるネットワークの実現」, NICT, 2011.
- [25]Proofpoint, “Proofpoint Targeted Attack Protection”, (http://www.proofpoint.com/jp/products/prod_tap.html) (2013/9).
- [26]中尾章宏, 「仮想化ノード特集」, 情報通信機構 NEWS No.6, 2010.
- [27]高野正次, 「映像配信サービスの発展に対応するトラフィック管理技術」, NTT ネットワーク基盤研究所, 2013.