

インターネットと計測技術 (第1回)

長 健二郎

株式会社インターネットイニシアティブ 技術研究所

1 はじめに

一般に計測はあらゆる技術にとって基礎であり基本です。我々は、日常生活の中でもさまざまな計測を行っています。例えば、体重計、血压計、万歩計、スピードメータなどの測定器が生活の中に入ってきて、誰もが簡単に測定をする事ができるようになりました。しかし、これらの測定器の仕組みは意外と知られていません。物さしで長さを計るように同じ単位系で直接測定できるものは少なく、ほとんどの測定はいわゆる間接測定をしています。また、普段なにげなく測っているものも、測定精度や誤差や結果のばらつきについてまじめに考え出すと途端に難しくなります。例えば、家庭用デジタル体重計の多くは、ひずみゲージで荷重によるゆがみを検知し電圧の変化として測定します。最近のデジタル体重計は 100g まで表示しますが、その精度は $\pm 200\text{g}$ 程度です。ちなみに、計測と測定という言葉は、ともに計量という意味で広く使われていますが、計測の方が測定手段の考案や結果の解析を含んだ広い概念を表します。

2 ネットワークデータの特徴とダイナミックス

ネットワークのデータは一般にバラツキが大きく、偏った分布を持つのが特徴です。これは、短時間にバースト的に転送を行なう構造や、少数の利用者が大半のトラフィックを占めるというような利用の偏りが原因にあります。例えばパケットサイズの分布を調べると、データを含まない TCP の ACK パケットの 40 バイトとイーサネットでの最大パケットサイズである 1500 バイトにピークがある事が知られています。また、トラフィック量やトポロジ等のネットワークデータには、出現頻度が大きさのべき乗に反比例するべき分布をはじ

め、さまざまな偏りを持つ分布が存在します。このように偏った分布を持つデータに対しては、単純な平均値はあまり意味を持ちません。

さらに、ネットワークでは、バグ、設定ミス、仕様の不整合、事故、メンテナンスなどによる異常が日常的に発生しています。そして、運用や機器設計では、日常的な異常を含めて正しく動作するよう考えなければなりません。同時に、このような異常を解析する事で、問題の発見やその原因の究明に繋がることもあります。したがって、目的に応じて、異常なデータをフィルタしたり、異常を含めて考えたり、異常に注目したりする訳です。

ネットワークプロトコルは階層に分けて設計されていますが、異なるレイヤに実装された機構は、それぞれが環境の変化に柔軟に適応するようになっています。輻輳制御に関連する技術を例にとると、イーサネットのコリジョン回避、インターフェイスでのパケットキューイング、TCP の輻輳制御、そして回線容量設計などは異なるレイヤで独立して機能します。また、それぞれの情報のフィードバック時間に応じて適応動作の時間粒度も違います。プロトコル仕様には、タイムアウトまでの時間など実装依存の部分も多く、実装の微妙な違いや、あるいは OS やハードウェアの違いからも、その組合せによって挙動の違いが出てくる事もしばしばです。

さらに、膨大なトラフィックが集約されるバックボーンでは、個別の挙動からは想像できないような全体の挙動が観測される場合もあります。インターネットは、無数の要素の相互作用の結果、全体としてみれば個別要素の総和以上の独立な振舞いをみせる複雑系の典型といえます。このようなインターネットの挙動は、実験室のネットワークでは容易に再現できません。したがって、生きたインターネットでの実測が欠かせないのです。なかには個別に見た時には些細な不備でも、

重なると大きな影響を持つ事があり、そのような原因の究明には、技術の詳細な知識も必要です。

3 ネットワーク計測の制約

実際のネットワークの計測にはさまざまな制約があり、その制約のもとで測定を行なうことになります。ネットワークは外部から観測するのが難しく、通常はネットワークの内部からの測定となるため、測定そのものが測定対象に影響を与えることを考慮しないといけません。

生きたネットワークを測定しようとする、実際にサービスしている機器にアクセスする必要がありますが、通常これらの機器へのアクセスは、一部の運用者へのみ許可されています。また、多くのネットワークは測定することを考慮して設計されていないので、測定のために構成変更が必要になったり、測定機器のスペースの確保などが発生するため、運用者の理解と協力が不可欠です。それでも実サービスに影響がでるような構成変更はなかなか出来ない、運用の現状を理解して実情にあった測定方法を工夫する事が重要です。

また、測定にはあまりコストをかけられない場合が多く、最新技術が導入されているルータを、それより性能が劣る計測機器、多くの場合は汎用 PC で測定するため得られるデータの精度にも限界があります。それでも多量のデータが収集できる場合には、統計処理によりある程度精度を上げることも可能です。

データの解析にあたっては、通信データの内容はプライバシーや企業機密に関わる場合もあるので取り扱いには十分な注意が必要です。データの中身だけでなく、IP アドレスや通信パターンからも通信相手や活動の概要が特定できる可能性があります。そのため、一般に測定データを組織外に持ち出す事は難しく、外部の研究者がデータを利用する障壁になっています。また、ネットワークデータの取り扱いに関しては法整備も整っていないのが現状です。

このような状況から、取得の難しいデータは貴重な技術資産となっていて、第三者が解析に使える汎用のデータを蓄積し公開する努力もなされています。そのいっぽう、汎用のデータは目的にあった精度や誤差を持つとは限らないので、利用するには注意が必要です。

4 インターネットの計測 – 掴みどころのないものを測る

ネットワークの状態を知ることはネットワークを運用するためだけでなく、プロトコルや機器の開発にも欠かせません。小規模なネットワーク運用のためなら、限定された範囲のネットワーク計測で済みますが、広域ネットワークの運用、グローバルなインターネットに関する取り決め、さまざまな場所で使われるネットワーク機器設計には、より一般化したデータを集める必要があります。例えば、輻輳によるパケットロスを低減するため、ルータのパッファサイズや回線容量の余裕設計を工学的に行なおうとすると、一般的なパケットの到着間隔やパケットサイズの分布を知る必要があります。

このように、インターネットにおける一般的な測定データというものが必要になってくるのですが、「インターネットにおける一般的な」という部分が曲者です。インターネットはさまざまなネットワークの集合体なので、測定する場所や時間によって異なる姿を見せます。ブロードバンドが普及した国とそうでない国、地域の言語や文化、企業と家庭などでインターネットの使い方にはずいぶん違いがあります。また、90 年代に World Wide Web が急速に広がり、2000 年以降ピアツーピア型のファイル共有、インターネット電話、ビデオ通信が増えてきたように、時代と共に使い方も変わってきています。

そこで、インターネットの一般性を求める事が重要なテーマとなります。しかし、インターネットは開いた系で、つねに変化、発展、拡大しています。自律分散型のインターネットには、中心もなければ代表点もなく、測る場所や時間によって違う姿が観測されます。すべてを網羅できない、たとえてきてもそれは過去のデータで現在の姿ではない訳ですから、答えのない問いへの取り組みです。この事は計測にとってやっかいです。真値がわからないと測定が正しいか検証のしようがないからです。しかし、現実にインターネットを運用する、あるいはプロトコルや機器を開発するためには、その時点で最善の一般性を模索し、将来予想を立て、また、それらを常に見直していく努力をするしかありません。

さらに、インターネットは大きく見ると技術面だけでなく、社会的、政策的、経済的な影響を受けていま

す。そして、インターネットを大局的に解析することは、社会、政策、経済なども考慮しないといけなくなります。

5 おわりに

ネットワークの計測は、見えないネットワークを見ようとする試みです。見えないネットワークだからこそ見える形にして把握する事が大切です。しかし、そのためには、ネットワーク技術やデータ解析の総合的な知識と理解が欠かせません。また、計測技術として見た場合には、インターネットの計測には、計測の面白さと難しさのさまざまな側面がこいま見えます。