NeTraMet による Root DNS サーバ群の計測

関谷 勇司 (sekiya@wide.ad.jp)

2005 年 2 月 2 日

概 要

WIDE Project における NeTraMet 計測結果の報告

1 NeTraMet とは

NeTraMet とは、CAIDA[1] によって開発された、トラフィックフローを計測するためのツールである。CAIDA は、主にインターネットにおける様々な計測に関する研究を行っているグループである。

NeTraMet は、RFC2720[2]、RFC2721[3]、RFC2722[4]、RFC2723[5]、RFC2724[6] にて標準化された、Real Time Flow Measurement(RTFM) に従って作成されている。フロー観測の記述言語として、Simple Ruleset Language(SRL) を用いており、柔軟なフロー計測を可能としている。

NeTraMet ツールは、実際にフロー計測を行う NeTraMet コマンドと、計測結果を SNMP にて取り出してファイルに記述する NeMaC コマンドの 2 つのコマンドから成る。Linux, FreeBSD, NetBSD といった一般的なフリー UNIX 系 OS の上で動作する。計測を行うためには、NeTraMet が動作するホストにて次の条件が必要となる。 (1) 計測対象となるトラフィックが、NeTraMet を動作させるインタフェースにて観測できる。 (2) 特権 (ルート権限) にて NeTraMet を動作させることができる。

すなわち、NeTraMet は実トラフィックを監視して計測するという静的な計測ツールであるため、スイッチングハブにおけるポートミラー設定やトラフィック分岐装置等によってトラフィックを複製し、ホストにて監視できる環境が必要となる。

NeTraMet パッケージの最新版は、http://www2.auckland.ac.nz/net/NeTraMet/から入手可能である。

2 NeTraMet による計測

mawi WG では、NeTraMet を使って Root DNS サーバ群への名前解決要求ならびに応答クエリの計測を行っている。これは、複数地点から Root DNS サーバ群への到達性を計測し、Root DNS サーバのエニーキャスティングの有効性を検証しようという活動の一部として行われている。 Root DNS サーバのエニーキャスティングとは、複数地点に同じ IP アドレスを持つ Root DNS サーバを設置して、同じアドレスブロックを BGP にて広告することによって、クエリの分散処理を行う技術である。これによって、RTT が小さくなり、サービス妨害攻撃や故障への耐性が増すと考えられている。

一方、NeTraMet の開発元である CAIDA においても、Root DNS サーバ群ならびに gTLD DNS サーバ群へのクエリの計測を行っている。この計測は CU Boulder, University Auckland, University of California San Diego(UCSD) の 3 地点で行われている。 計測結果は論文 [7, 8] に示されている。

CAIDA による最新の計測結果は、http://www.caida.org/cgi-bin/dns_perf/main.pl にて確認できる。

3 WIDE Project における計測

前説にて述べたとおり、WIDE Project においても NeTraMet による Root DNS サーバへのクエリ計測を行っている。設置地点は、慶応大学と東京大学の 2 地点である。慶応大学においては、慶応大学湘南藤沢キャンパスと、そのネットワークの上流となる WIDE Project との中間に位置するスイッチにおいてポートミラー設定を行い、NeTraMet を設置した。東京大学においては、東京大学とその上流となる学術情報ネッ

トワーク (SINET) との間において光分岐装置を設置し、NeTraMet を設置した。

すなわち、慶応大学においては湘南藤沢キャンパスにて発生する Root DNS サーバへのクエリを計測し、東京大学においては東京大学全体から発生する Root DNS サーバへのクエリを計測することとなる。

計測するためのホストの仕様は、次の通りである。慶応大学は CPU に Pentium III 800Mhz、メモリ 256MB を搭載したホストに、1000base-SX インタフェースを用いて計測を行う。東京大学では Xeon 3Ghz 2 個、メモリ 1GB のホストに、同じく 1000base-SX インタフェースを用いて計測を行う。なお、NeTraMet パッケージのバージョンはどちらも Ver 5.1b2 を利用している。これは、Ver 4 以前の NeTraMet では、DNS クエリの RTT を正確に測定することができないというバグがあったからである。

計測対象は A.root-servers.net から M.root-servers.net までの 13 個の Root DNS サーバである。計測に利用した SRL を図 1 に示す。

この SRL 設定によって、各 Root DNS サーバへの問い合わせならびに応答を記録し、それぞれの数や応答を得られるまでにかかった Round Trip Time(RTT) を記録している。この計測結果は NeMaC によって 5 分単位で記録される。

NeMaC によって記録されるデータの例をは図 2 に示す。 この図では見やすいように適切に改行を入れた。#Format で始まるコメントの段落に、NeMaC が記録する数値の意味が明記されている。19 という数値にて始まるそれぞれの段落が、一つの Root DNS サーバに関する 5 分間の計測結果を示している。

WIDE Project においては、この NeTraMet による Root DNS サーバ群へのクエリ計測を 2003 年 9 月から試験的に開始し、2004 年 1 月から 2 地点における本格的な計測を開始した。

4 計測結果

NeTraMet による測定結果の例として、図 3 に 2004 年 8 月 7 日の慶応大学における測定結果を示す。

また、図 4 に、2004 年 12 月 31 日の慶応大学における測定結果を示す。

縦軸が RTT(ms) もしくは Root DNS サーバに向けて出された名前解決要求のクエリ数を示し、横軸が時間を示している。グラフ中の「＋」の点が RTT を示し、「×」の点がクエリ数を示す。

どちらの日ににおいても、m.root-servers.net に対する RTT が最小となっている。これは m.root-servers.net は WIDE Project によって運営されており、WIDE Project のネットワークから近い場所に位置しているためである。

注目すべきなのは、i.root-servers.net に対する結果である。2004 年 8 月 7 日と 2004 年 12 月 31 日の i.root-servers.net に関する結果のみを図 5 に示す。8 月と 12 月の結果で RTT が大きく異なっている。これは、i.root-servers.net が分散拠点を東京に設置し、エニーキャスト [9] を利用したサービスを開始したためである。このため、8 月の時点では海外にある i.root-servers.net のホストに送られていたクエリが、東京の分散拠点にて処理されるようになったため、RTT が小さくなったと考えられる。

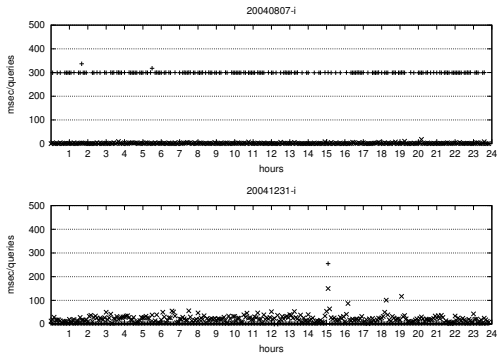

図 5: i.root-servers.net に対する計測結果の比較

さらに、東京大学における計測結果を図 6 に示す。

慶応大学に比べ、東京大学の方が観測されるクエリの全体数が多い。そのため、慶応大学の測定結果に比べて、m.root-servers.net に多量のクエリが送信されていることがはっきりと判別できる。

WIDE Project における NeTraMet の計測結果は、http://dnstap.nc.u-tokyo.ac.jp/NeTraMet/にて公開している。

5 結論

このように、NeTraMet を利用して Root DNS サーバへのクエリを計測することができる。しかし、今回の計測にてわかったこととして、次の問題点があげられる。

- NeTraMet 設置ならびに運用コスト

  NeTraMet を設置し運用するためには、ポートミラーやトラフィック分岐装置等によってトラフィックを複製し、監視する必要があるため、それ相応の設置コスト、運用コストが必要となる。

- エニーキャスト

  多くの分散拠点によってエニーキャスティングが行われた場合、RTT やクエリ数の変化はつかむことができるが、どの分散拠点にて処理されているのかをつかむことが難しい。

- NeTraMet のバッファ不足

  NeTraMet が SRL による柔軟なフロー処理を行うため、単にダンプを行って監視する場合に比べ、処理が重くなる傾向にある。また、複数のインタフェースを監視対象とした場合や、たくさんのトラフィックが発生した場合、フローを記憶しておくためのバッファが不足し、正確に記録できない場合が発生する。今回の計測においても、東京大学での計測においてバッファ不足が発生した期間があった。NeTraMet のログによると、2004 年の 5 月から 12 月まで不定期に発生していた。NeTraMet プロセスに割り当てられるメモリ量の上限を増やし、NeTraMet プロセス起動時に -t オプションにてバッファサイズを増加させることによって対処した。

参考文献

[1] Cooperative Association for Internet Data Analysis(CAIDA), http://www.caida.org/

[2] Traffic Flow Measurement: Meter MIB, N. Brownlee, RFC2720, October 1999

[3] RTFM: Applicability Statement, N. Brownlee, RFC2721, October 1999

[4] Traffic Flow Measurement: Architecture, N. Brownlee, C. Mills and G Ruth, RFC2722, October 1999

[5] SRL: A Language for Describing Traffic Flows and Specifying Actions for Flow Groups, N. Brownlee, RFC2723, October 1999

[6] RTFM: New Attributes for Traffic Flow Measurement, S. Handelman, S. Stibler, N. Brownlee and G. Ruth, RFC2724, October 1999

[7] DNS Root/gTLD Performance Measurements, Nevil Brownlee, kc claffy and Evi Nemeth, USENIX LISA2001 Conference, December 2001

[8] Response time distributions for global name servers, Nevil Brownlee and Ilze Ziedins, PAM2002, March 2002

[9] Distributing Authoritative Name Servers via Shared Unicast Addresses, T. Hardie, RFC3258, April 2002

```
define DNS = 53;
define PP_ICMP_ECHO     =   1;

define PP_UDP_DNS       =  11;

define PP_TCP           = 192;
define PP_OK_SYNACK     =   1;  #   ->SYN,  <-SYN+ACK pairs
define PP_OK_SYNRST     =   2;  #   ->SYN,  <-SYN+RST pairs

define PP_OK_MULTI      =   8;  #   ->DATA, <-ACK for more than one packet
define PP_OK_SINGLE     =  16;  #   ->DATA, <-ACK 'lone' packet
define PP_OK_INGROUP    =  32;  #   ->DATA, <-ACK single packet in a group

define A_ROOT = 198.41.0.4/32;       #  Verisign, Dulles, Va
define B_ROOT = 192.228.79.201/32;   #  ISI, Marina del Ray, Ca
define C_ROOT = 192.33.4.12/32;      #  Cogent, Herndon, Va
define D_ROOT = 128.8.10.90/32;      #  U Maryland, Md
define E_ROOT = 192.203.230.10/32;   #  NASA Ames, Ca
define F_ROOT = 192.5.5.241/32;      #  ISC, Palo Alto, Ca
define G_ROOT = 192.112.36.4/32;     #  DoD NIC, Vienna, Va
define H_ROOT = 128.63.2.53/32;      #  ARL, Abdereen, Md
define I_ROOT = 192.36.148.17/32;    #  KTH, Stockholm
define J_ROOT = 192.58.128.30/32;    #  Verisign, Dulles, Va
define K_ROOT = 193.0.14.129/32;     #  RIPE NCC, Amsterdam
define L_ROOT = 198.32.64.12/32;     #  IANA, Los Angeles, Ca
define M_ROOT = 202.12.27.33/32;     #  WIDE, Tokyo

define B_ROOT_OLD = 128.9.0.107/32;      #  ISI, USC, Ca
define J_ROOT_OLD = 198.41.0.10/32;      #  NSI, Herndon, Va

define TestDestAddress =
   if DestPeerAddress == A_ROOT
      { store FlowKind :=  1\;  store FlowClass := 0\; }
   else if DestPeerAddress == B_ROOT || DestPeerAddress == B_ROOT_OLD
      { store FlowKind :=  2\;  store FlowClass := 0\; }
   else if DestPeerAddress == C_ROOT
      { store FlowKind :=  3\;  store FlowClass := 0\; }
   else if DestPeerAddress == D_ROOT
      { store FlowKind :=  4\;  store FlowClass := 0\; }
   else if DestPeerAddress == E_ROOT
      { store FlowKind :=  5\;  store FlowClass := 0\; }
   else if DestPeerAddress == F_ROOT
      { store FlowKind :=  6\;  store FlowClass := 0\; }
   else if DestPeerAddress == G_ROOT
      { store FlowKind :=  7\;  store FlowClass := 0\; }
   else if DestPeerAddress == H_ROOT
      { store FlowKind :=  8\;  store FlowClass := 0\; }
   else if DestPeerAddress == I_ROOT
      { store FlowKind :=  9\;  store FlowClass := 0\; }
   else if DestPeerAddress == J_ROOT || DestPeerAddress == J_ROOT_OLD
      { store FlowKind := 10\;  store FlowClass := 0\; }
   else if DestPeerAddress == K_ROOT
      { store FlowKind := 11\;  store FlowClass := 0\; }
   else if DestPeerAddress == L_ROOT
      { store FlowKind := 12\;  store FlowClass := 0\; }
   else if DestPeerAddress == M_ROOT
      { store FlowKind := 13\;  store FlowClass := 0\; }


optimise 3;

   if SourcePeerType == IPv4 save;
   else ignore;

   if SourceTransType == UDP save;
   else ignore;

   TestDestAddress;  # Sets FlowKind
   if FlowKind == 0 nomatch;
   else {

      if DestTransAddress == DNS save;  # Avoid 'match on non_DNS flow' msg
      else ignore;

      save ToTurnaroundTime = 120.11.0!0 & 4.2.10!7000;
      count;
      }

set dns_root_wide;
statistics;
format
  FlowRuleSet FlowIndex FirstTime SourcePeerType SourceTransType
  "  " FlowKind FlowClass
  "  " ToPDUs FromPDUs
  "  " ToLostPDUs FromLostPDUs
  "  (" ToTurnaroundTime
  ")";
```

図 1: SRL 設定

```
#flows starting at 15:00:00 Mon 3 Jan 2005 (UTC)

#Format: flowruleset flowindex firsttime sourcepeertype
#sourcetranstype flowkind flowclass topdus frompdus tolostpdus
#fromlostpdus (toturnaroundtime1) Time:

19 4 1039 1 17 2 0 6253468 5318492 1152512 217536 (5 2 10 7000 63 11 0
0 1913 1904 1902 1900 1839 1903 1903 1902 1902 1874 1917 1914 1843
1834 1903 1828 1862 1830 1830 1903 1904 1903 1835 1834 1830 1829 1829
1829 1903 1900 1904 1902 1834 1915 1904 1903 1900 1899 1903 1899 1933
1925 1905 1829 1904 1907 1908 1907 1922 1904 1904 1902 1904 1905 1834
1838 1914 1868 1858 1827 1913 1837 1915)

19 5 1040 1 17 9 1 12735380 12663671 678923 607214 (5 2 10 7000 63 11
0 0 3025 2976 3055 3017 3048 3045 3081 2978 3011 3011 3010 3055 3063
3071 3056 2967 3013 3057 3040 3022 2943 3035 3020 3056 2941 3040 3053
3030 3030 3041 3026 3055 3027 3058 3046 3063 3030 3013 3042 3041 3023
3036 3028 3045 3046 3061 3076 3053 3055 3042 3025 2940 3047 3038 3051
3063 2988 2942 3065 3025 3030 3017 3048)

19 6 1041 1 17 3 1 16703780 16469505 923629 689354 (6 2 1965 7000 120
11 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 3 1 1 1 0 2 0 1 0 2 2 2 1 1 2 2 2
3 0 2 1 1 1 2 2 1 1 0 2 1 1 1 0 2 1 0 0 0 3 2 0 2 0 1 5 0 0 2 5 2 5 4
2 2 2 3 4 1 3 3 1 3 4 6 1 2 4 3 3 5 0 5 6 0 2 3 2 2 3 2 2 1 1 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)

19 7 1046 1 17 13 0 37799759 36803333 1532996 536570 (6 2 7874 7000
120 11 0 0 648 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)

19 8 1046 1 17 6 1 13515243 13363098 757920 605775 (5 2 10 7000 75 11
0 0 2651 2647 2510 2494 2486 2484 2484 2515 2589 2562 2550 2566 2566
2576 2586 2571 2562 2532 2550 2564 2552 2574 2587 2551 2545 2547 2569
2544 2587 2551 2549 2674 2636 2655 2655 2646 2646 2659 2641 2528 2554
2586 2536 2564 2580 2575 2486 2471 2495 2490 2495 2504 2505 2662 2660
2529 2575 2594 2595 2472 2518 2495 2486 2555 2571 2579 2580 2586 2577
2561 2575 2566 2550 2527 2522)

19 9 1049 1 17 11 1 13839538 13330392 1106929 597783 (5 2 10 7000 38
11 0 0 2499 2507 2614 2571 2485 2612 2500 2498 2496 2537 2534 2500
2509 2520 2483 2511 2531 2529 2537 2494 2504 2511 2542 2566 2574 2575
2481 2611 2565 2604 2605 2476 2524 2591 2592 2600 2506 2516)

19 10 1049 1 17 5 1 13773863 13606764 755134 588035 (5 2 10 7000 51 11
0 0 2679 3025 2702 2699 2680 2712 2785 2651 2679 2674 2731 2692 2654
2722 2721 2694 2779 2779 2737 2566 2690 2686 2699 2819 2754 2820 2694
2675 2722 2692 2647 2646 2646 2716 2692 2687 2670 2670 2669 2735 2749
2747 2639 2637 2659 2692 2749 2632 2699 2697 2645)

19 11 1055 1 17 8 1 10480084 10353402 692623 565941 (5 2 10 7000 96 11
0 0 2116 2028 2074 2088 1932 1945 1953 1957 1969 1957 1982 1983 1983
1998 2035 2040 2055 2084 2121 2089 2004 2015 2012 2070 2076 2077 2008
2014 2043 2079 2099 2060 2009 2048 2022 2060 2091 2070 2101 2125 2084
2090 2108 2109 2114 2019 2003 2079 2029 1999 2003 2030 2015 2018 2030
2002 2034 2054 2075 2094 2116 2010 2009 2034 2039 2027 2055 2073 2111
2114 2090 2062 2067 2074 2055 2015 2023 2044 2043 2025 2098 2035 2014
2014 2030 2003 2076 2084 2005 2017 2042 2052 2010 2013 2054 2050)

19 12 1073 1 17 13 1 58832809 57840355 2323094 1330640 (5 2 10 7000 75
11 0 0 2754 2739 2879 2836 2575 2574 2573 2571 2800 2840 2749 2852
2884 2882 2882 2752 2885 2931 2932 2632 2626 2595 2589 2584 2579 2827
2735 2732 2712 2721 2908 2961 2834 2750 2750 2614 2614 2621 2579 2595
2594 2592 2591 2792 2907 2775 2711 2872 2851 2911 2892 2952 2571 2632
2596 2610 2619 2619 2582 2581 2571 2571 2626 2624 2604 2632 2617 2604
2600 2571 2500 2550 2544 3013 2835)

19 13 1082 1 17 3 0 6711146 4381917 2520040 190811 (5 2 10 7000 64 11
0 0 2143 2099 2047 2095 2143 2124 2100 2098 2095 2090 2158 2158 2164
2168 2094 2112 2196 2144 2131 2154 2133 2047 2076 2090 2135 2094 2178
2131 2089 2093 2094 2097 2179 2094 2100 2144 2070 2074 2081 2071 2080
2064 2099 2071 2065 2074 2100 2072 2124 2068 2083 2047 2144 2068 2065
2070 2133 2149 2191 2175 2130 2048 2125 2145)

19 14 1082 1 17 1 0 6782804 6413979 671041 302216 (5 2 10 7000 76 11 0
0 2079 2062 2079 2083 2080 2078 2079 2079 2079 2070 1995 2058 2079
2083 2081 2081 2083 2079 2080 2079 2082 2060 2059 2063 2062 2061 2063
2060 2060 2060 2061 2078 2078 2079 2081 2080 2080 2081 2079 2080 2079
1919 2080 2080 2069 2083 2085 1998 2082 2082 2083 2081 2082 2083 2081
2080 2081 2082 2082 2082 2084 1997 2080 2079 2082 2069 2082 2082 1997
2080 2062 2063 2073 2078 2059 1997)

19 15 1083 1 17 12 1 16445117 16233637 901572 690098 (5 2 10 7000 57
11 0 0 2128 2161 2168 2144 2129 2158 2115 2223 2239 2160 2143 2208
2225 2125 2128 2155 2141 2149 2155 2112 2124 2118 2118 2118 2153 2161
2225 2126 2156 2163 2158 2111 2136 2164 2131 2144 2120 2136 2171 2195
2229 2135 2160 2190 2229 2114 2148 2140 2139 2153 2129 2150 2143 2153
2129 2138 2159)

19 16 1121 1 17 4 1 14702378 14518514 817334 633470 (5 2 10 7000 80 11
0 0 2341 2306 2244 2263 2340 2293 2296 2295 2356 2356 2360 2394 2380
2329 2329 2363 2255 2306 2243 2243 2274 2275 2299 2295 2211 2204 2358
2354 2295 2284 2281 2229 2208 2198 2243 2293 2306 2181 2278 2370 2244
2354 2349 2275 2259 2216 2183 2310 2221 2308 2244 2290 2284 2285 2285
2286 2363 2199 2306 2320 2319 2353 2359 2320 2338 2321 2320 2320 2341
2321 2369 2344 2320 2349 2320 2361 2351 2281 2283 2271)
```

図 2: NeMaC 記録結果

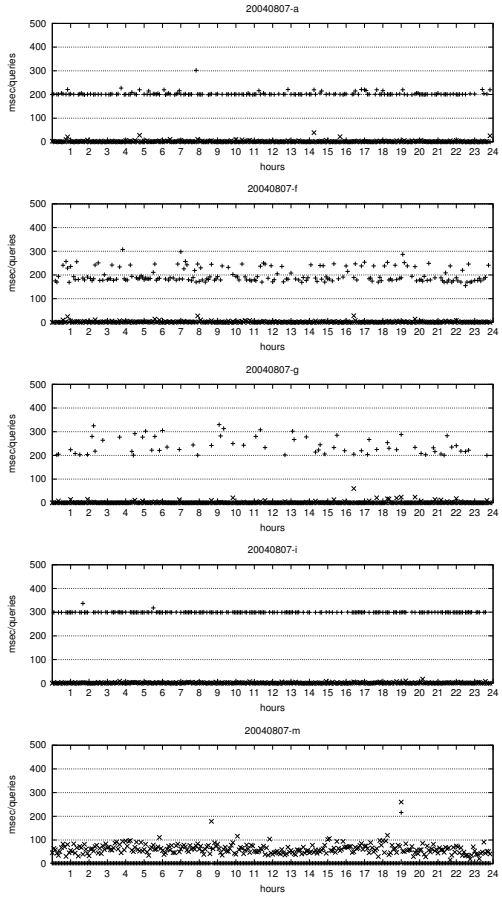

図 3: 2004 年 8 月 7 日：慶応大学

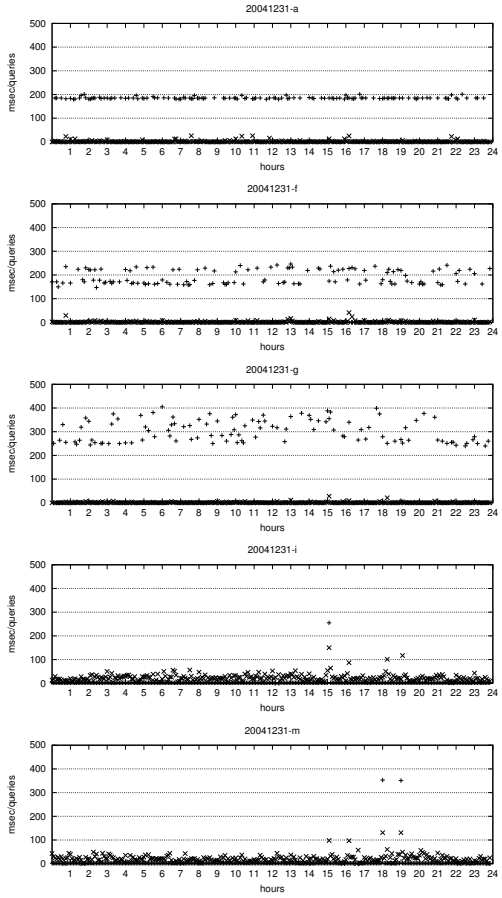

図 4: 2004 年 12 月 31 日：慶応大学

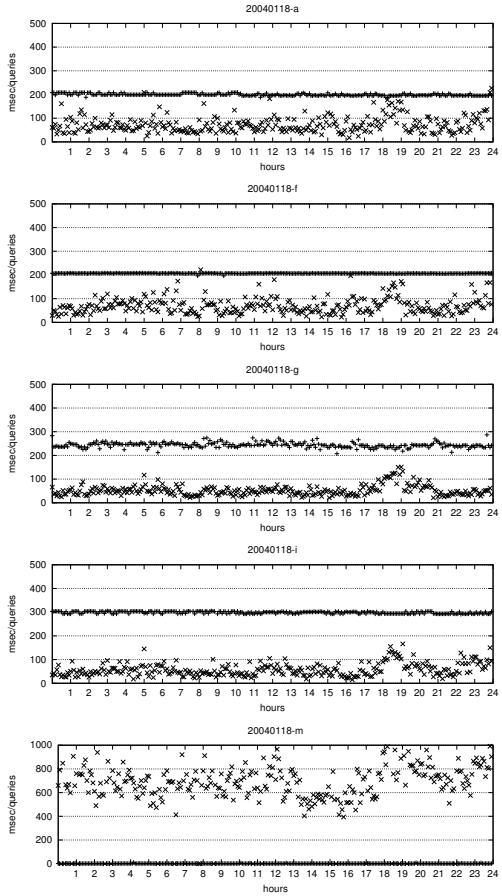

図 6: 2004 年 1 月 18 日：東京大学