

# Impact of Traffic Mix and Packet Sampling on Anomaly Visibility

Bernhard Tellenbach, Daniela Brauckhoff, Martin May

TIK Report 275

Department of Information Technology and Electrical Engineering

Swiss Federal Institute of Technology (ETH)

Zurich, Switzerland

{brauckhoff, may, tellenbach}@tik.ee.ethz.ch}

## Abstract

Packet sampling methods such as Cisco’s NetFlow are widely employed by large scale networks to reduce the amount of traffic data measured. A key problem with packet sampling is that it is inherently a lossy process, discarding potentially useful information. In this paper, we empirically evaluate the impact of traffic mix and packet sampling on anomaly visibility using traffic traces collected at four different border routers of a medium scale national ISP. These traffic traces consist of *unsampled* flow traces collected during the Blaster and Witty worm outbreak. We use our knowledge of the Blaster and Witty anomaly to establish a baseline of normal traffic against which we measured the size of the anomaly at various sampling rates. We analyze the traffic mix characteristics of the baseline traffic and use this knowledge to evaluate its impact on anomaly visibility. Our results confirm previous findings suggesting that entropy metrics are more resilient to packet sampling than volume metrics. But surprisingly, we find that simple metrics like unique destination port counts can be superior to entropy metrics even for sampling rates of 1 out of 10000. Furthermore, we find that traffic mix characteristics can compensate or even boost anomaly visibility in sampled views for sampling rates up to 1 out of 10000.

## 1 Introduction

Fast and accurate detection of network traffic anomalies is a key factor in providing a reliable and stable network infrastructure. With efficient anomaly detection systems, large scale network infrastructure providers are able to limit the impact of traffic anomalies by applying traffic filtering at the border of their network. But, while most of the anomaly detection methods have been designed under the assumption that they have a complete view on the traffic flowing through a sensor, this assumption does not hold when we consider the vast amount of traffic data flowing through the border routers

of large scale networks. As a result, instead of accounting for each arriving packet at each router, traffic sampling has emerged as the dominant means to summarize the vast amount of traffic data continuously collected by providers of large scale networks. While being attractive because of efficiency and availability, sampling is inherently a lossy process, where many packets are discarded without inspection. Thus, sampled traffic is an incomplete and more importantly, a biased approximation of the underlying traffic trace, as small flows are likely to be missed entirely.

Packet sampling is now being discussed for many years [8, 9, 15], however, the following two questions remained largely unanswered: (i) are there anomaly detection metrics that can preserve anomalies in sampled view better than others; and (ii) how do traffic mix characteristics impact anomaly detection on sampled views? While question (ii) is still unanswered, we addressed question (i) in [1] where we empirically evaluated the impact of packet sampling on anomaly detection metrics for the blaster worm outbreak. By using a set of flow traces collected at a single border router of a national ISP during the outbreak of the blaster worm, we found that entropy metrics are more resilient to packet sampling than other metrics like e.g., flow counts. A similar finding was published by Mai *et al* [11] in parallel to our work. While this was a first and important step in finding an answer to question (i), it was left to future research whether these results suggest that it is necessary to calculate complex entropy metrics, or if simpler metrics like e.g., the total number of unique destination IPs observed, are equally well-suited for anomaly detection at high sampling rates, even if they are more severely affected by sampling. And even more important, it remained unclear if and how these findings depend on the traffic mix seen at a specific sensor.

In this paper, we rely on two unique week-long datasets of *unsampled* flow records from all border routers of a national ISP to (1) empirically evaluate the impact of packet sampling and traffic mix characteristics on anomaly detection metrics and (2) to identify met-

rics that are equally well-suited for anomaly detection at high sampling rates, even if they are more severely affected by sampling.

Both evaluations involved the following four steps:

1. simulating packet sampling to construct sampled views of the traffic traces
2. building the ideal baseline by separating anomalous traffic from normal traffic
3. calculation of anomaly detection metrics for the baseline and for all traffic
4. study the impact of sampling by measuring and comparing anomaly visibility at decreasing sampling rates

In a last step, we identified the traffic mix characteristics of the baseline traffic. This latter one was mainly a requirement of the traffic mix impact analysis.

Both evaluations provide surprising results: First, we present evidence that a suitable traffic mix can negate the destructive effects of packet sampling on anomaly visibility up to sampling rates of 1 out of 10000. Furthermore, we find that simple count metrics like the number of unique source ports per interval can be superior to complex entropy metrics even if they are more affected by sampling. In particular, the visibility of the Witty worm, when measured in unique destination port counts, is significantly higher than the corresponding flow destination port entropy.

The remainder of this paper is organized as follows. We next provide an overview of related work. Then in section 2, we introduce in detail our data set and methodology. In Section 3, we discuss our study of the traffic mix characteristics from our four routers. Section 4 presents the impact of packet sampling and traffic mix characteristics on anomaly visibility. Finally, we conclude and outline directions for future work in Section 5.

## 2 Related Work

The most prevalent and widely-deployed method of sampling traffic is *packet sampling*, where a router inspects every  $n$ -th packet (uniformly at random), and records its features (addresses, ports, protocol, and flags). Packet sampling is attractive because it is computationally efficient, requiring minimal state and counters, and is implemented in high-end routers today (e.g., with NetFlow [3]). As such, many large networks (ISPs and enterprizes) are now using packet sampling to obtain rich views of traffic directly from routers. But despite its benefits when it comes to reducing the vast amount of traffic to an amount that can be handled and stored in a convenient way, it does no longer provide a complete view to an anomaly detection system. Previous work has largely focused on analyzing this bias, devising better sampling strategies [2], and recovering

statistics (moments and distribution) of the underlying traffic trace using inference [4–6]. Furthermore, sampled traffic views have recently been used for anomaly detection with considerable success [8, 9]. But, little is known about the fidelity of the sampled stream for these applications. Two notable studies are [11] and [10]. In [11], Mai *et al* analyzed how packet sampling impacts three specific portscan detection methods, TRWSYN [7], TAPS [13] and entropy-based profiling method of [9, 15]. Recently, this work was extended to analyze the impact of other sampling schemes in [10]. Both studies conclude that packet sampling is inadequate to detect anomalies using these detection methods. However, in [1] we followed a more general approach by empirically evaluating the impact of packet sampling on anomaly detection metrics rather than specific detection methods. Using flow traces collected during the blaster worm outbreak, we showed that entropy metrics are less affected by sampling and expose the blaster worm even at high sampling rates. This study was a first important step toward answering question (i). A first step toward answering question (ii) is [12]. The authors of [12] looked at the propagation of several anomalies from networks with different traffic mix characteristics. They based their analysis on sampled views from routers of two large scale networks. Their results suggest that traffic mix characteristics can be an important factor. But because of the lack of traffic mix characteristics of the unsampled traffic, a detailed impact analysis was not possible.

## 3 Methodology

In this Section, we present the measurement data and methodology for packet sampling used in this study. Moreover, we introduce the metrics we compare and the methodology we apply for measuring anomaly visibility.

### 3.1 Measurement Data

For this present study, we use two week-long extracts from a comprehensive 3-year data set of *unsampled* NetFlow trace database from the Swiss Education and Research Network (SWITCH), a medium-sized Swiss backbone network. The traces have been collected at the ISP’s four border gateway routers and therefore provide us with a complete view of all Internet traffic that enters and leaves the ISP network. Furthermore, because the roles of the border routers are quite different - private peering with an international research network only, peering with international carriers only, or combinations of the two (see figure 1) - the traffic mix seen by each of them does differ significantly and show different characteristics.

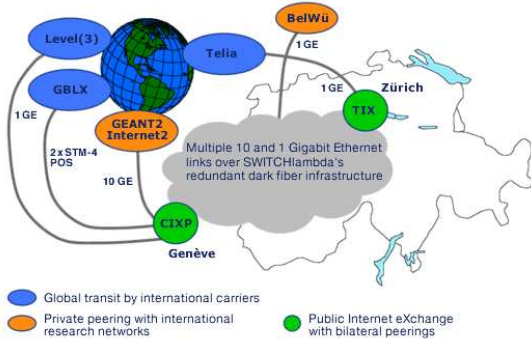


Figure 1: Global connectivity of the Swiss Education and Research Network (SWITCH)

The first week-long extract, which is already used in our previous work, was collected between the 8th and 15th of August, 2003 and contains the well-known *Blaster* worm. Additionally, we study a second week-long data set which was collected from 17-21, March, 2004 during the outbreak of the *Witty* worm. The two datasets allows us to analyze two well-known scanning-based anomalies with diverse characteristics. In particular, those characteristics are:

- Witty uses UDP random scanning while Blaster does TCP random scanning for target identification,
- Witty infected only about 15'000 hosts (running a specific firewall product) while Blaster infected between 200'000 and 500'000 hosts worldwide,
- Witty uses a fixed source port (4000) and variable destination port while Blaster uses a variable source port and fixed destination port (135).

Hence, the different properties provide an insight into the impact of anomaly diversity with regard to the transport protocol used, the impact on network traffic, as well as the spreading strategy into account.

### 3.2 Packet Sampling on Flow Data

The sampling method applied in this study is random probabilistic packet sampling. That is, when sampling at a rate of  $q = 0.01$  (or 1 in 100), we select each packet with a probability of  $q$  or discard it with a probability of  $1 - q$ . Random probabilistic packet sampling is the most widely applied sampling method today due to its small computational overhead.

For being able to apply packet sampling to a Net-Flow trace, we need to identify individual packets in that trace, including each packet's arrival time and size in bytes. To obtain this information from the flow trace, we calculate the packet size by dividing the flow size by the number of bytes in that flow. The time stamp of the packet is randomly selected within the flow bounds with a resolution of milliseconds. For a detailed discussion

on this topic, please refer to [1].

### 3.3 Considered Detection Metrics

We compare a total of 15 metrics which are frequently applied for anomaly detection in backbone networks. These are:

- volume metrics (number of bytes, packet, and flows)
- feature entropy based on flows and packets for destination and source IP as well as destination and source port.

Additionally, we want to assess whether the complex entropy metrics are the only metrics well-suited for anomaly detection using sampled trace datasets. Specifically, we examine if simple feature counts such as the total number of unique destination IPs observed are equally suitable for anomaly detection. Thus, we introduce four additional metrics which we did not include in our previous study:

- unique source IP counts, destination IP counts,
- source port counts, destination port counts

Note that the approach of comparing multiple metrics allows us to draw general conclusions, rather than focusing on a particular anomaly detection method.

### 3.4 Measuring Anomaly Visibility

Our approach to determine the traffic baseline and to measure the anomaly visibility is illustrated in Fig. 2. We determine the baseline for each metric and sampling rate by applying some simple heuristics. For Blaster that is, we remove all TCP packets from the measured trace which have a destination of 135, and a size of either 40, 44, or 48 bytes. For Witty we remove all UDP packets matching the following heuristic: Packet size between 796 and 1307 bytes, and source port of 4000.

We measure anomaly visibility  $vis$  as the normalized Euclidean distance between two measurement points, in the baseline  $\hat{x}$  and in the total traffic  $x$

$$vis = (x - \hat{x}) / \hat{x} \quad (1)$$

In Fig. 2(a) for example, the visibility in the first interval after the initial Blaster outbreak at 17:30 (UTC) for the *unsampled* flow count metric is computed as  $(2'000'000 - 1'000'000) / 1'000'000$ , and equals to 1. That is, the anomaly has the same size as the pure baseline traffic and is very well visible.

## 4 Traffic Mix and Detection Metrics

In this section, we discuss the composition of the traffic traces we used throughout this analysis and introduce

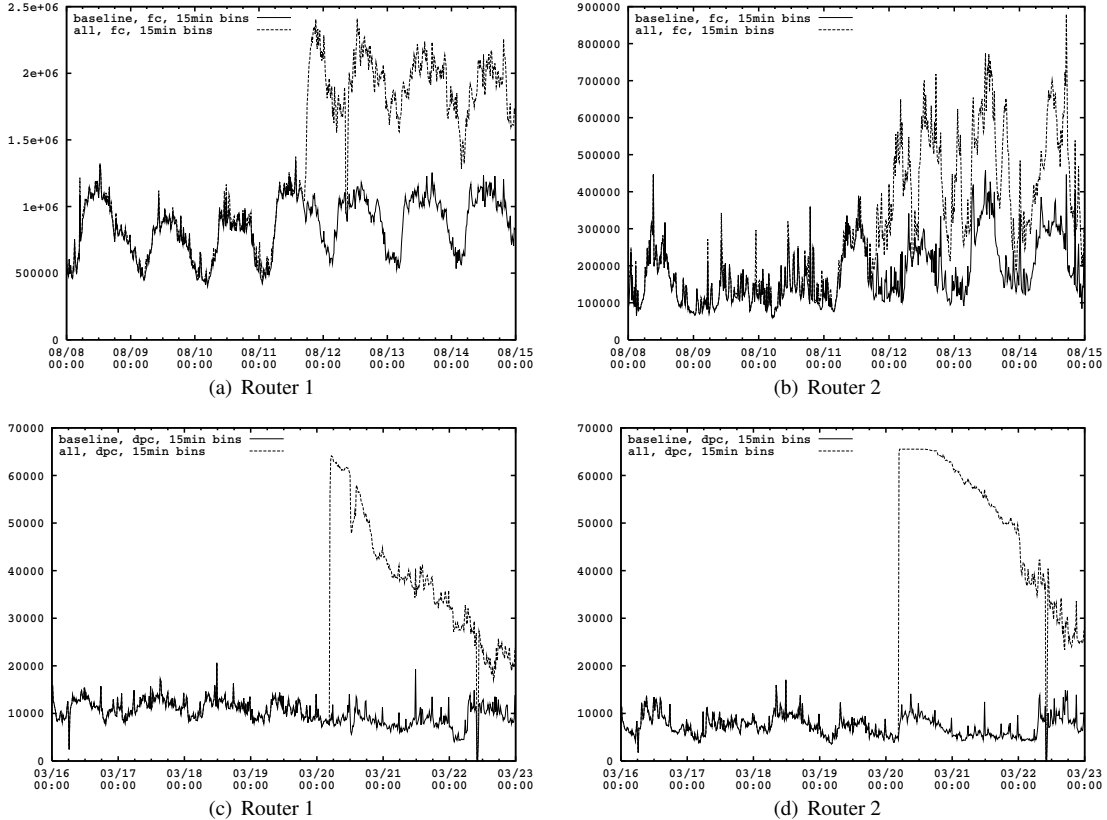


Figure 2: (a) and (b): Number of flows per interval on router 1 and 2 for baseline and total traffic during the outbreak of the blaster worm. No sampling. (c) and (d): Number of destination ports per interval on router 1 and 2 for baseline and total traffic during the outbreak of the witty worm. No sampling.

the metrics applied to the Blaster Worm traces and the traces containing the Witty Worm outbreak.

Our starting point was the observation that the traffic traces from our four routers show very different characteristics. To obtain a more detailed view on these characteristics, we analyzed the traces using the same set of metrics as described in [1]. For each of those metrics, we determined the average and the 95 percentile value (or standard deviation) on an hour basis. In order to limit the impact of isolated fluctuations, we based our per hour calculations on a total of five hours of the five preceding workdays.

Furthermore, we introduce two additional metrics that will help to understand some of the effects observed when analyzing the impact of traffic mix and packet sampling: the average number of packets per flow and average number of packets per IP address. Detailed results can be found in our technical report [14]. In our next step, we used results from our unsampled traces to identify the metrics that highlight best the Blaster and/or Witty anomaly.

## 4.1 Metrics for the Blaster Worm

Figure 3 shows the visibility of six different metrics on all four routers for the week containing the Blaster outbreak. According to Figure 3, destination IP address count/entropy, flow count and flow destination port entropy show significant exposure of the Blaster anomaly. From these metrics we selected those showing the impact of traffic mix on packet sampling best: flow counts and flow destination IP address entropy. All plots that were relevant for this selection can be found in our technical report [14]. Surprisingly, Figure 3 reveals "Blaster activity" already prior to the outbreak of the blaster worm. Possible root causes of this observation are: (1) our heuristic to determine the baseline is not accurate enough, (2) there was a significant amount of pre-outbreak blaster (scan-) activity. (1) is supported by the fact that our heuristic catches a limited amount of "blaster-activity" even some weeks before the outbreak. (2) is supported by the fact that the visibility of the destination IP address count (see Figure 3) has a ramp-like increase toward the actual outbreak of the blaster worm on 11th of August 16:35 (UTC) on all four routers. To investigate this in more detail, we might re-

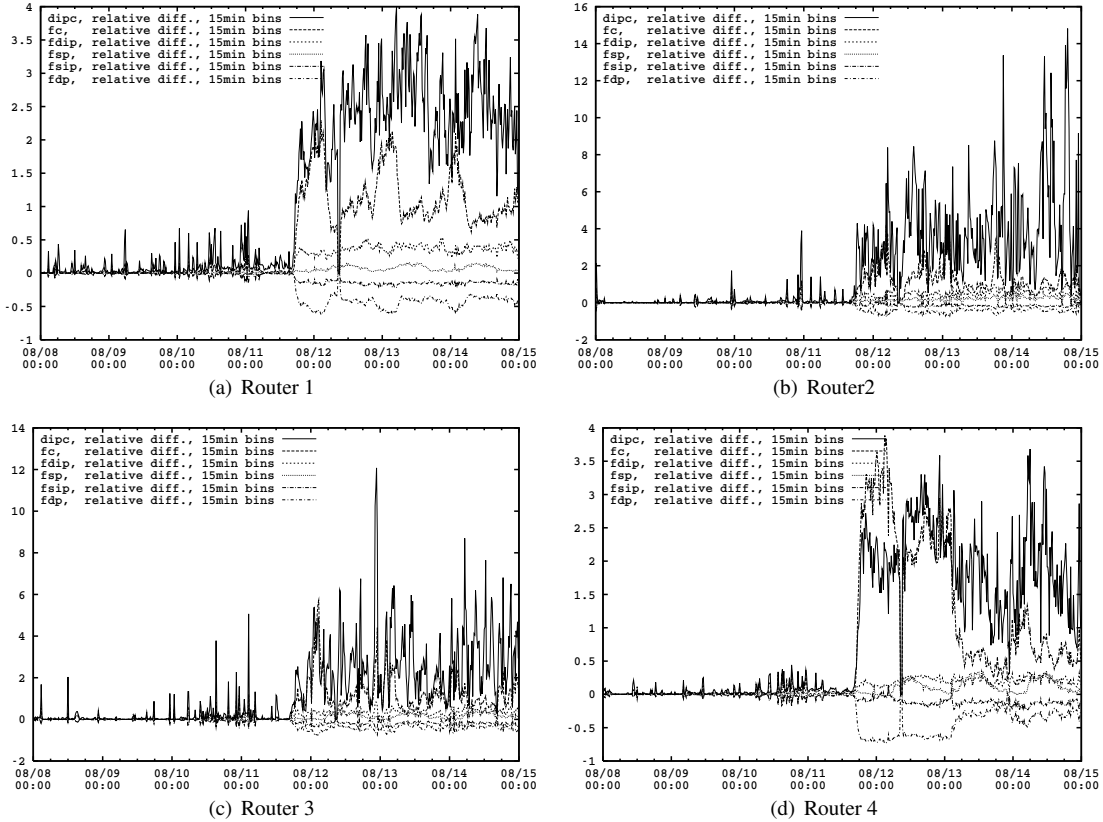


Figure 3: Relative difference of (cdip) destination ip address count, (fc) flow count, (fdip) flow destination ip address entropy, (fsp) flow source port entropy, (fsip) flow source ip address entropy and (fdp) flow destination port entropy metric on our four border routers vs. Date and Time (UTC). No sampling.

fine our heuristic by e.g. counting only blaster flows from hosts that scan more than one host in our network. We plan to investigate this in future work.

## 4.2 Metrics for the Witty Worm

By analyzing the relative distance plots for the Witty traces, we identified the following set of metrics with good anomaly visibility: destination IP address count, flow destination port entropy, flow count and destination port count. But this time, we time we did not have traffic mix impact analysis in mind but a comparison of simple count metrics to more complex entropy metrics. A careful analysis of all of the anomaly visibility plots (see [14]) showed that flow source and destination port entropy and unique source and destination port counts are best suited for this purpose. Furthermore, it is important to note that in contrast to our heuristic for filtering Blaster activity, our Witty heuristic appears to be perfect: The anomaly visibility plots (see Figure 5) do not show any Witty activity prior to its outbreak.

## 5 Anomaly Visibility in Sampled Traffic

When moving from anomaly visibility in unsampled traffic to anomaly visibility in sampled traffic, there are several factors that are relevant to the change in anomaly visibility. Specifically, those are the:

- sampling type (and rate),
- anomaly traffic characteristics,
- baseline traffic characteristics, and
- anomaly metric

From those factors, we first study the impact of baseline traffic characteristics on different anomaly detection metrics when random packet sampling is applied at different rates. The basis for this study is the trace dataset containing the Blaster worm for which we already have a thorough understanding from our previous work. Furthermore, we successfully validated the results from our previous study [1] applying the same methodology to the Witty worm anomaly. Due to space restrictions, we published the relevant results separately in [14].

Finally, we assess the quality of complex entropy metrics with regard to more simple feature counts like

total number of unique destination IPs observed. We do this by comparing the visibility for four count metrics (unique source IP addresses, destination IP addresses, source ports, destination ports) to the visibility of the corresponding entropy metrics.

### 5.1 Impact of Traffic Mix

We study the impact of backbone traffic mix on sampled detection metrics by having a closer look at router 2 and 4. Figure 4 shows the relative difference for flow count and flow destination IP address entropy for these selected routers.

A comparison of the flow count plots (on the left side) reveals a similar impact of packet sampling on both routers: The anomaly visibility is decreased significantly when going from unsampled traffic to traffic sampled at a rate of 1:100. However, the Blaster visibility on router 4 is approximately twice as strong as on router 2. This indicates that router 4 has received much more Blaster traffic than router 2. Looking at the traffic mix, we found that the average number of packets per flow in the baseline is significantly higher than in the anomalous traffic (consisting of one- or two-packet scans) for both routers. Consequently, the chances that an anomalous flow is selected (i.e., a minimum of one packet of that flow is kept) are much lower than for baseline flows, and flow counts are equally affected on both routers.

Comparing the flow destination IP address entropy metric (on the right side) reveals a significant difference: Packet sampling *increases* Blaster visibility on router 4, while it *decreases* the visibility on router 2. An increase in visibility for sampled traffic is counter-intuitive at first.

Hour	packets/flow		packets/dst IP	
	R 2	R 4	R 2	R 4
16:00	34±10	16±2	137±61	46±13
17:00	36±11	16±3	142±82	45±15
18:00	43±13	16±3	240±121	48±20
19:00	38±8	14±3	216±188	40±15
20:00	40±9	14±2	133±66	38±14

Table 1: Characteristics of the baseline of router 2 and 4: Average and standard deviation for selected hours over the past five workdays.

However, a close look at the baseline traffic mix characteristics for router 2 and 4 (see Table 1) provides us with an explanation: The average number of baseline packets per destination IP address on router 2 is at least three times as high as on router 4. This implies that the chance of discarding a specific IP address in the baseline is significantly higher on router 4 than on router 2. Therefore, the flow destination IP address entropy

decreases faster with increased sampling rates for router 4. As seen in Figure 4 (left hand), this impact of packet sampling and traffic mix is strong enough to amplify the anomaly on one router while decreasing the same metric on the other router with higher sampling rates. We found similar effects for several other metrics.

### 5.2 Feature Entropy vs. Unique Count Metrics

Before assessing whether costly entropy metrics are worth calculating, or if simple unique feature counts are equally well-suited for anomaly detection, let us briefly discuss how Witty and Blaster affect the source and destination port distribution metrics. For Witty, we observe a concentration of the distribution (on the fixed source port 4000) and a dispersion of the destination port distribution (towards a uniform distribution) due to random destination port scanning activity. The Blaster worm leads to an opposite effect, the destination port distribution gets more concentrated (on fixed destination port 135), while the source port distribution gets more dispersed (due to randomly selected source ports). Additionally, on some routers that see traffic only on a small number of source ports, an increased number of source ports is observed.

There are basically two criteria for our assessment of entropy and count metrics: (1) anomaly visibility, and (2) stability of the metric for higher sampling rates. We illustrate our argumentation with Fig. 5 where we plot the flow entropy for the source and destination port, as well as the unique count of these features for the Witty worm.

In Fig. 5(a) and 5(c), the relative distance for the destination port entropy and counts is shown for different sampling rates. Comparing the two Figures, we observe that the destination port count metric shows a larger anomaly distance over all sampling rates. Although, the relative distance for the destination port entropy is less affected by packet sampling, it stays below the distance of the count metric.

However, the situation regarding the source ports is different. In Fig. 5(b) and 5(d), we observe that the source port entropy metric shows a large anomaly distance, while the source port count shows a very small relative distance. This is due to the fact that anomalies which cause a concentration of a feature distribution are not captured in a simple count metric.

In summary, we conclude that simple count metrics are better suited for detecting anomalies that cause a broadening of a feature distribution, e.g., scanning activity. However, anomalies that have an impact on the feature distribution without affecting the total number of unique events (such as dispersion of the distribution or concentration of the feature distribution on one port

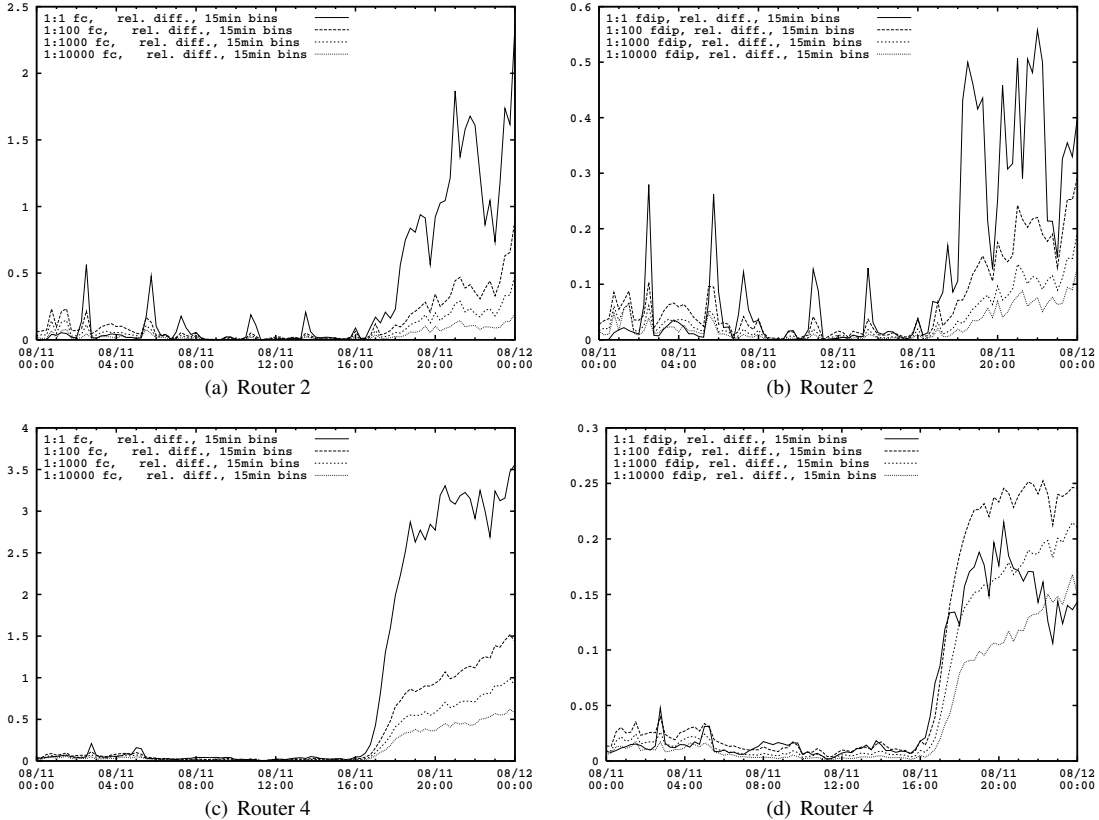


Figure 4: Impact of traffic mix and packet sampling on flow count (fc) and flow destination IP address entropy (fdip) metrics during the outbreak of the blaster worm.

or IP address) are not visible in simple count metrics. For detecting this type of anomalies, entropy is our metric of choice. Moreover, we have shown that the impact of packet sampling on entropy and count metrics depends largely on the background traffic mix, e.g., the flow length distribution, as well as on the type of anomaly.

## 6 Conclusion

In this paper, we empirically evaluate the impact of traffic mix and packet sampling on anomaly visibility. Starting with two week-long datasets of unsampled traffic records from four border routers, we first ask how traffic mix affects anomaly metrics in combination with packet sampling. To answer this question we apply our unique and general methodology of anomaly visibility – which treats anomalies as deviation from an idealized baseline – to a set of 15 metrics that are frequently used in anomaly detection. This approach allows us to draw general conclusions, rather than focusing on a particular anomaly detection method.

By comparing the 15 metrics for the four border gateway routers at different sampling rates, we find that the

traffic mix characteristics of the baseline (and the anomaly) have a large impact on anomaly visibility at high sampling rates. Moreover, we find that visibility of certain metrics, e.g., flow destination IP entropy, is even more pronounced by packet sampling, and elaborate on the reasons for this interesting finding. We conclude that, depending on the traffic mix characteristics and the anomaly under consideration, it might still be convenient to use sampling rates up to 1:10<sup>0</sup>000. Having found that distribution-based metrics like entropy are resilient to packet sampling, we compare anomaly visibility in computationally rather complex entropy metrics and simple feature counts considering only the width of a feature distribution. We find that both metrics have their pros and cons: While simple feature counts are well-suited for detection of scanning-based anomalies that induce a widening of the port or IP distribution, entropy metrics are better suited for detecting anomalies that induce a concentration of the distribution on a specific port or IP address.

The results presented in this paper open up new directions for research on devising detection metrics or on understanding the impact of traffic mix and/or packet sampling for different types of anomalies.

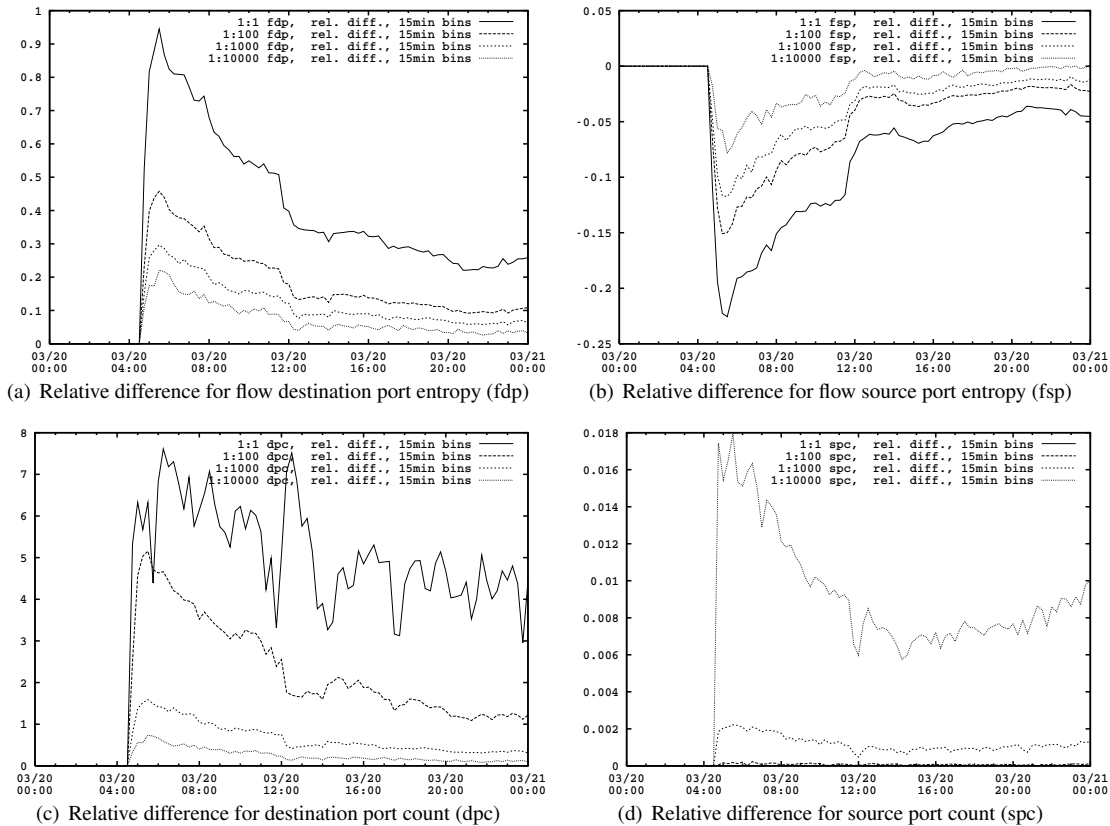


Figure 5: Impact of packet sampling on selected anomaly detection metrics during the outbreak of the witty worm for increasing sampling rates (all Router 1).

## References

- [1] BRAUCKHOFF, D., TELLENBACH, B., WAGNER, A., MAY, M., AND LAKHINA, A. Impact of packet sampling on anomaly detection metrics. In *IMC '06: Proceedings of the 6th ACM SIGCOMM on Internet measurement* (New York, NY, USA, 2006), ACM Press, pp. 159–164.
- [2] CHOI, B.-Y., PARK, J., AND ZHANG, Z.-L. Adaptive random sampling for total load estimation. In *In proceedings of the IEEE International Conference on Communications (ICC '03)* (2003).
- [3] Cisco NetFlow. At [www.cisco.com/warp/public/732/Tech/netflow/](http://www.cisco.com/warp/public/732/Tech/netflow/).
- [4] DUFFIELD, N., LUND, C., AND THORUP, M. Properties and prediction of flow statistics from sampled packet streams. In *ACM SIGCOMM Internet Measurement Workshop* (2002).
- [5] DUFFIELD, N., LUND, C., AND THORUP, M. Estimating Flow Distributions from Sampled Flow Statistics. In *ACM SIGCOMM* (Karlsruhe, August 2003).
- [6] HOHN, N., AND VEITCH, D. Inverting Sampled Traffic. In *Internet Measurement Conference* (Miami, October 2003).
- [7] JUNG, J., PAXSON, V., BERGER, A., AND BALAKRISHNAN, H. Fast portscan detection using sequential hypothesis testing, 2004.
- [8] KIM, M.-S., KANG, H.-J., HUNG, S.-C., CHUNG, S.-H., AND HONG, J. W. A Flow-based Method for Abnormal Network Traffic Detection. In *IEEE/IFIP Network Operations and Management Symposium* (Seoul, April 2004).
- [9] LAKHINA, A., CROVELLA, M., AND DIOT, C. Mining Anomalies Using Traffic Feature Distributions. In *ACM SIGCOMM* (Philadelphia, August 2005).
- [10] MAI, J., CHUAH, C.-N., SRIDHARAN, A., YE, T., AND ZANG, H. Is sampled data sufficient for anomaly detection? In *IMC 2006 (To appear)* (Rio de Janeiro, Brazil, October 2006).
- [11] MAI, J., SRIDHARAN, A., CHUAH, C.-N., ZANG, H., AND YE, T. Impact of packet sampling on portscan detection. *IEEE Journal on Selected Areas of Communications* 24, 12 (2006), 2285–2298.
- [12] SOULE, A., RINGBERG, H., SILVEIRA, F., REXFORD, J., AND DIOT, C. Detectability of traffic anomalies in two adjacent networks. In *Passive and Active Measurement Conference (PAM)* (Louvain-la-Neuve, Belgium, 2007).
- [13] SRIDHARAN, A., YE, T., AND BHATTACHARYYA, S. Connectionless port scan detection on the backbone. In *Malware workshop, held in conjunction with IPCCC* (Phoenix, AZ, April 2006).
- [14] TELLENBACH, B., BRAUCKHOFF, D., AND MAY, M. Impact of traffic mix and packet sampling on anomaly visibility, technical report. Tech. rep., ETH Zurich, April 2007.
- [15] XU, K., ZHANG, Z.-L., AND BHATTACHARYYA, S. Profiling internet backbone traffic: Behavior models and applications. In *ACM Sigcomm 2005* (Philadelphia, PA, August 2005).





# A Traffic Mix Characteristics of the Blaster baseline

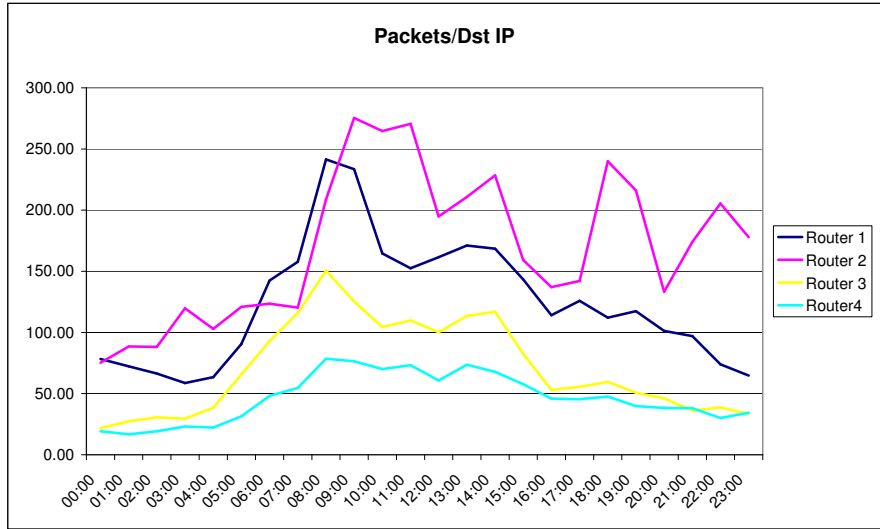
## BASELINE DATA STATISTICS (whole week)

	Router 1	Router 2	Router 3	Router 4
<b>Bytes (15 min)</b>				
AVG	1.91E+10	7.92E+09	2.12E+09	9.79E+09
STDEV	7.38E+09	4.21E+09	1.76E+09	6.45E+09
STDEV/AVG	38.67%	53.16%	83.12%	65.88%
<b>Packet Count (15 min)</b>				
AVG	2.15E+07	6.16E+06	1.82E+06	8.24E+06
STDEV	5.55E+06	2.27E+06	1.11E+06	3.75E+06
STDEV/AVG	25.73%	36.79%	60.79%	45.47%
<b>Flows (15 min), IDs</b>				
AVG	8.59E+05	1.81E+05	1.25E+05	5.96E+05
STDEV	2.13E+05	8.23E+04	5.99E+04	1.85E+05
STDEV/AVG	24.80%	45.48%	47.85%	31.03%
<b>Flow DesIP Ent. BL</b>				
AVG	1.15E+01	1.03E+01	1.09E+01	1.45E+01
STDEV	1.20E+00	1.86E+00	1.72E+00	1.31E+00
STDEV/AVG	10.42%	18.03%	15.87%	9.06%
<b>Flow Dest IP Count BL</b>				
AVG	1.99E+05	5.50E+04	4.10E+04	2.41E+05
STDEV	6.52E+04	3.58E+04	2.57E+04	9.21E+04
STDEV/AVG	32.70%	65.04%	62.63%	38.22%

## BASELINE DATA STATISTICS (workdays, per hour)

Hour	FlowSize											
	Router 1			Router 2			Router 3			Router 4		
	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev
00:00	21119.95	7158.24	33.89%	34183.84	14286.13	41.79%	8747.34	3913.00	44.73%	8027.51	2396.70	29.86%
01:00	19502.48	4251.26	21.80%	36335.60	16729.16	46.04%	9663.32	3766.31	38.98%	6774.75	1932.19	28.52%
02:00	18666.79	3224.60	17.27%	35727.23	13338.82	37.34%	8399.87	2521.55	30.02%	7183.11	2695.13	37.52%
03:00	17962.18	3554.50	19.79%	39868.10	11860.68	29.75%	7315.00	2505.72	34.25%	9575.97	2521.22	26.33%
04:00	19268.90	3286.94	17.06%	37871.69	14855.29	39.23%	10667.33	6566.49	61.56%	9073.41	4179.86	46.07%
05:00	21002.76	3026.78	14.41%	34569.20	11544.00	33.39%	20072.14	14578.68	72.63%	11865.37	4592.44	38.64%
06:00	24315.00	2284.31	9.39%	34837.82	11546.30	33.14%	19303.13	2697.43	13.97%	18394.10	4904.42	26.66%
07:00	26387.28	2930.77	11.11%	40118.67	10316.71	25.72%	22854.62	4546.10	19.89%	22302.74	6620.19	29.68%
08:00	28407.99	3562.41	12.54%	43343.88	9768.46	22.54%	24914.08	5245.79	21.06%	25322.67	6648.81	26.26%
09:00	28330.69	3002.00	10.60%	44428.35	10334.63	23.26%	23522.28	5720.26	24.32%	25945.63	4734.60	18.25%
10:00	27124.38	3499.88	12.90%	40098.20	11563.95	28.84%	21741.58	4977.14	22.89%	24793.21	4750.24	19.16%
11:00	25766.69	3771.79	14.64%	49604.34	13778.81	27.78%	22984.93	4365.66	18.98%	25602.21	4396.13	17.17%
12:00	28634.98	3463.35	12.09%	48350.43	13571.25	28.07%	24450.45	6343.61	25.94%	26146.95	4546.31	17.39%
13:00	28247.62	3784.28	13.40%	48630.63	11458.93	23.56%	24642.40	3590.47	14.57%	27338.99	5112.80	18.70%
14:00	27472.96	2751.56	10.02%	47686.02	8349.56	17.51%	25501.91	4732.38	18.56%	25973.57	4553.50	17.53%
15:00	24631.90	2814.10	11.42%	43742.39	12923.92	29.55%	21669.42	4384.74	20.23%	23470.64	2920.19	12.44%
16:00	21833.96	2734.22	12.52%	45603.17	17445.58	38.26%	16817.90	4249.38	25.27%	20133.20	2902.57	14.42%
17:00	20559.58	2247.38	10.93%	48196.22	15097.69	31.33%	19052.74	8272.51	43.42%	18986.25	3042.44	16.02%
18:00	20690.80	2363.31	11.42%	56601.50	25463.31	43.45%	16586.90	5546.00	33.44%	17349.51	2104.45	12.13%
19:00	18737.84	2351.39	12.55%	49913.88	13724.46	27.50%	14640.27	2569.18	17.55%	15403.56	2310.55	13.83%
20:00	17874.33	2553.80	14.29%	53943.93	19198.60	35.59%	14343.92	6448.58	44.96%	15140.22	2766.16	18.27%
21:00	18065.28	2131.13	11.80%	50106.92	19402.06	38.72%	11599.39	5242.34	45.19%	14336.05	1993.62	13.91%
22:00	17055.06	2478.00	14.53%	58109.14	19047.91	32.78%	12141.26	4587.37	37.78%	11131.11	2603.22	23.39%
23:00	17180.79	2691.12	15.66%	49153.67	24819.35	50.43%	13006.46	6681.41	51.37%	12404.27	4659.48	37.56%
TOTAL	22451.67	3163.21	14.42%	44709.37	14601.07	32.73%	17276.61	5168.75	32.57%	17612.29	3737.80	23.32%

Hour	cnt_byt											
	Router 1			Router 2			Router 3			Router 4		
	Avg	95 percentil	avg_dev 95	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev
00:00	1.265E+10	1.798E+10	42.22%	4.877E+09	7.217E+09	47.98%	7.874E+08	1.837E+09	133.29%	4.148E+09	5.383E+09	29.79%
01:00	1.114E+10	1.491E+10	33.82%	4.939E+09	8.457E+09	71.24%	6.205E+08	6.650E+08	39.39%	3.490E+09	3.935E+09	12.75%
02:00	1.059E+10	1.296E+10	22.37%	4.574E+09	7.049E+09	54.11%	4.726E+08	6.729E+08	42.40%	3.462E+09	4.730E+09	36.63%
03:00	1.041E+10	1.267E+10	21.70%	4.299E+09	7.011E+09	63.10%	4.464E+08	6.005E+08	34.53%	3.336E+09	5.623E+09	42.84%
04:00	1.187E+10	1.393E+10	17.33%	4.706E+09	7.374E+09	56.70%	6.183E+08	8.258E+08	33.55%	3.839E+09	5.900E+09	53.68%
05:00	1.640E+10	1.994E+10	21.59%	5.098E+09	7.071E+09	38.70%	1.619E+09	3.117E+09	92.53%	5.895E+09	8.322E+09	41.18%
06:00	2.385E+10	2.682E+10	12.47%	7.460E+09	1.074E+10	43.94%	2.727E+09	4.296E+09	57.53%	1.150E+10	1.423E+10	23.67%
07:00	2.782E+10	3.208E+10	15.32%	1.051E+10	1.408E+10	33.92%	4.138E+09	5.508E+09	33.10%	1.591E+10	1.863E+10	17.10%
08:00	2.970E+10	3.377E+10	13.70%	1.225E+10	1.652E+10	34.87%	4.944E+09	7.755E+09	56.86%	1.777E+10	2.123E+10	19.46%
09:00	3.002E+10	3.318E+10	10.52%	1.257E+10	1.646E+10	30.90%	4.916E+09	6.033E+09	22.73%	1.955E+10	2.217E+10	13.42%
10:00	2.962E+10	3.468E+10	17.06%	1.052E+10	1.492E+10	41.77%	4.218E+09	5.190E+09	23.03%	1.761E+10	1.998E+10	13.45%
11:00	2.885E+10	3.296E+10	14.22%	1.377E+10	2.241E+10	62.73%	4.821E+09	5.973E+09	23.90%	1.967E+10	2.376E+10	20.79%
12:00	3.136E+10	3.616E+10	15.33%	1.527E+10	2.702E+10	76.96%	5.505E+09	6.688E+09	21.50%	2.207E+10	2.525E+10	14.42%
13:00	3.098E+10	3.468E+10	11.93%	1.491E+10	2.357E+10	58.07%	5.158E+09	6.704E+09	29.98%	2.270E+10	2.835E+10	24.90%
14:00	3.046E+10	3.325E+10	9.14%	1.332E+10	1.986E+10	49.05%	4.946E+09	6.442E+09	30.23%	2.116E+10	2.617E+10	23.67%
15:00	2.656E+10	3.022E+10	13.80%	1.008E+10	1.561E+10	54.87%	3.642E+09	4.385E+09	20.42%	1.699E+10	1.996E+10	17.58%
16:00	2.331E+10	2.820E+10	20.96%	9.143E+09	1.286E+10	40.61%	2.570E+09	3.497E+09	36.07%	1.294E+10	1.562E+10	20.71%
17:00	1.994E+10	2.250E+10	12.88%	8.697E+09	1.188E+10	36.59%	2.043E+09	2.720E+09	33.15%	1.064E+10	1.389E+10	30.48%
18:00	1.965E+10	2.232E+10	13.59%	8.690E+09	1.366E+10	57.13%	1.773E+09	2.106E+09	18.77%	9.202E+09	1.306E+10	41.97%
19:00	1.783E+10	2.006E+10	12.45%	8.010E+09	1.188E+10	48.34%	1.421E+09	1.734E+09	22.04%	8.022E+09	1.032E+10	28.65%
20:00	1.722E+10	1.996E+10	15.93%	8.788E+09	1.482E+10	68.61%	1.497E+09	2.450E+09	63.62%	7.710E+09	9.857E+09	27.84%
21:00	1.611E+10	1.828E+10	13.45%	7.023E+09	1.144E+10	62.94%	1.225E+09	1.923E+09	56.96%	6.699E+09	9.194E+09	37.24%
22:00	1.406E+10	1.665E+10	18.45%	6.938E+09	1.081E+10	55.84%	1.044E+09	1.259E+09	20.63%	5.511E+09	6.430E+09	16.67%
23:00	1.264E+10	1.411E+10	11.62%	6.106E+09	1.059E+10	73.50%	9.949E+08	1.518E+09	52.56%	4.845E+09	5.749E+09	18.64%
TOTAL	2.096E+10	2.426E+10	17.16%	8.857E+09	1.347E+10	52.60%	2.59E+09	3.50E+09	41.62%	1.15E+10	1.41E+10	26.15%



BASELINE DATA STATISTICS (workdays, per hour)

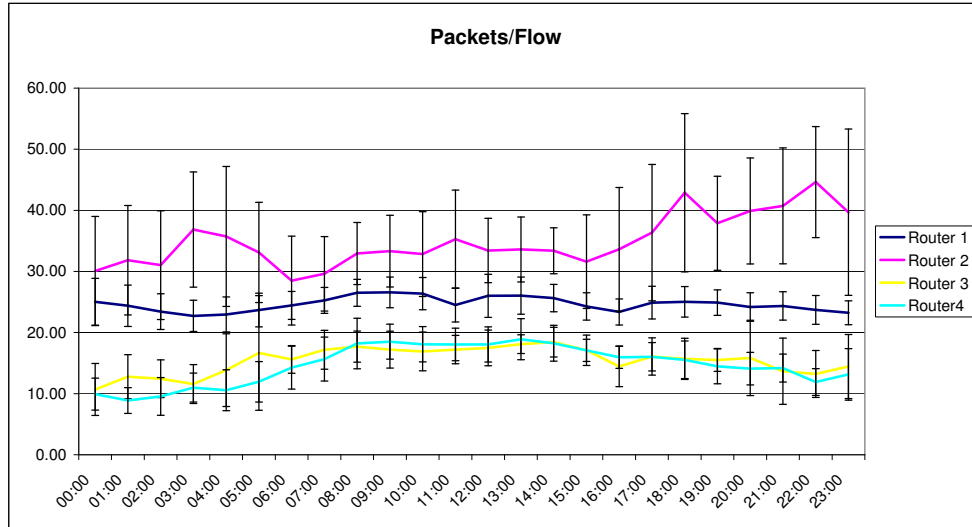
Hour	Packets/Dst IP											
	R94			Router 2			Router 3			Router 4		
	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev
00:00	78.32	32.01	40.87%	75.27	49.99	66.41%	21.83	15.52	71.06%	19.23	7.66	39.83%
01:00	72.32	28.12	38.88%	88.51	73.54	83.08%	27.26	10.41	38.20%	16.71	5.85	35.02%
02:00	66.45	22.11	33.27%	88.14	68.97	78.25%	30.62	13.40	43.75%	19.21	10.12	52.66%
03:00	58.64	12.50	21.31%	119.76	56.44	47.13%	29.40	17.77	60.44%	23.03	6.76	29.34%
04:00	63.43	13.36	21.06%	102.88	41.73	40.56%	38.52	25.47	66.13%	22.25	9.63	43.29%
05:00	90.55	29.75	32.86%	120.90	52.18	43.16%	65.66	53.99	82.24%	31.43	16.96	53.96%
06:00	142.48	48.50	34.04%	123.50	60.34	48.86%	92.91	45.18	48.63%	48.18	26.14	54.26%
07:00	157.73	34.56	21.91%	120.37	50.41	41.88%	115.95	59.34	51.18%	54.66	26.34	48.19%
08:00	241.50	209.86	86.90%	208.76	179.84	86.15%	150.63	116.25	77.18%	78.57	42.91	54.61%
09:00	233.51	141.12	60.43%	275.44	184.48	66.98%	125.35	74.04	59.07%	76.40	40.62	53.16%
10:00	164.64	59.46	36.12%	264.68	135.71	51.27%	104.41	56.24	53.86%	70.06	30.86	44.04%
11:00	152.48	53.31	34.96%	270.58	116.26	42.97%	109.84	35.63	32.44%	73.29	31.06	42.38%
12:00	161.50	48.93	30.30%	194.90	82.63	42.40%	100.22	48.89	48.78%	60.81	21.62	35.54%
13:00	171.11	67.97	39.72%	210.71	162.13	76.94%	113.51	53.88	47.47%	73.54	34.36	46.72%
14:00	168.49	81.58	48.41%	228.41	146.66	64.21%	116.96	48.73	41.66%	67.81	28.70	42.33%
15:00	143.35	44.22	30.85%	159.30	67.75	42.53%	82.50	40.77	49.42%	57.65	20.64	35.81%
16:00	114.12	43.59	38.19%	136.99	60.86	44.43%	53.00	31.13	58.73%	45.97	12.95	28.16%
17:00	125.90	42.81	34.01%	142.16	82.71	58.18%	55.59	27.80	50.01%	45.42	14.52	31.97%
18:00	112.04	34.44	30.74%	239.91	121.14	50.50%	59.55	40.58	68.14%	47.59	19.67	41.33%
19:00	117.43	40.59	34.57%	216.11	187.52	86.77%	50.73	21.11	41.61%	39.79	14.93	37.53%
20:00	101.27	28.04	27.69%	133.28	65.75	49.34%	46.21	26.56	57.47%	38.25	14.46	37.79%
21:00	97.06	29.04	29.92%	173.99	140.05	80.49%	36.10	26.75	74.09%	38.00	13.37	35.20%
22:00	73.92	12.93	17.49%	205.47	139.73	68.00%	38.70	29.86	77.17%	29.97	13.72	45.78%
23:00	64.76	10.85	16.76%	177.89	157.80	88.71%	33.45	17.38	51.97%	34.43	19.41	56.37%
TOTAL	123.87	48.74	35.05%	169.91	103.53	60.38%	70.79	39.03	56.28%	46.34	20.14	42.72%







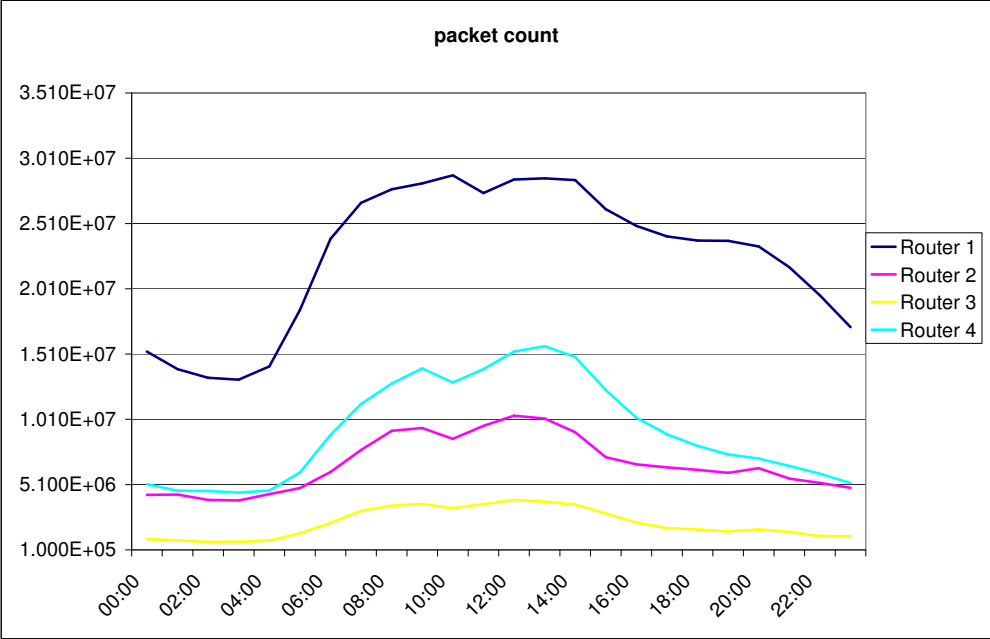
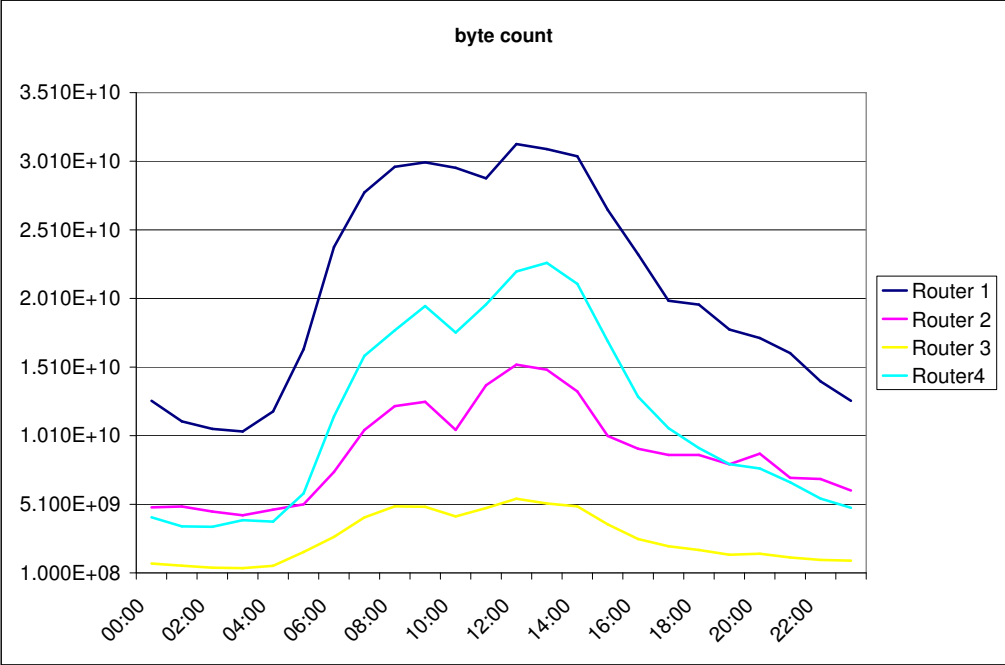




**BASELINE DATA STATISTICS (workdays, per hour)**

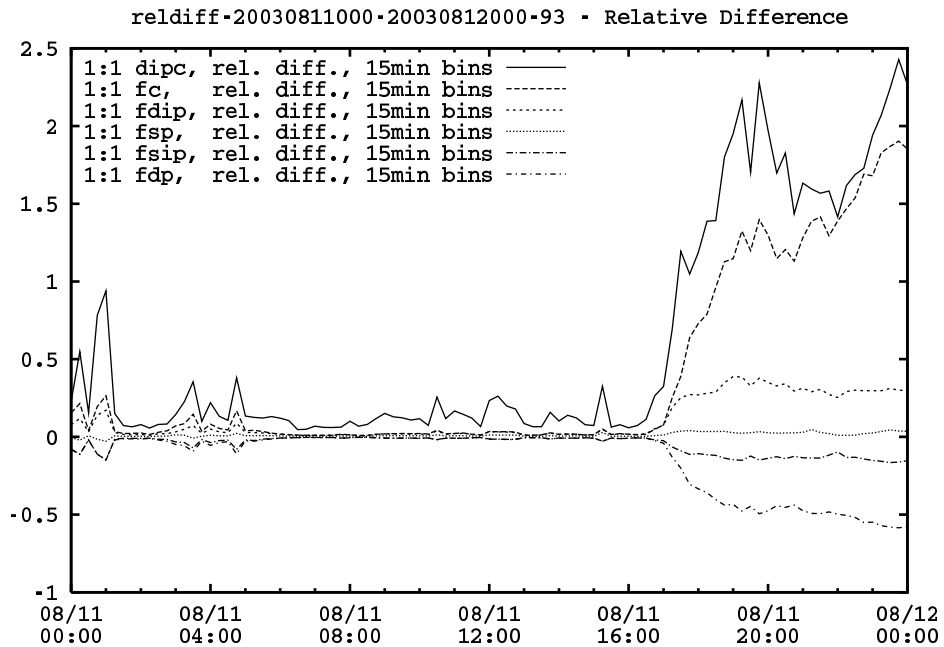
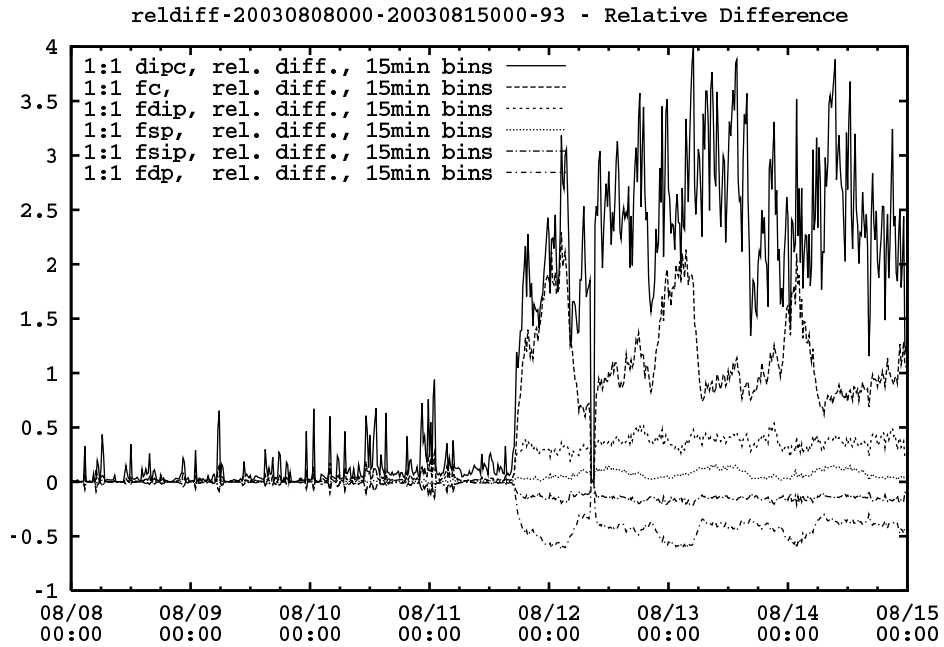
Hour	Packets/Flow											
	Router 1			Router 2			Router 3			Router 4		
	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev	Avg	StDev	Rel.StDev
00:00	25.03	3.83	15.28%	30.08	8.93	29.69%	10.68	4.25	39.75%	9.92	2.61	26.33%
01:00	24.39	3.39	13.89%	31.83	8.96	28.15%	12.77	3.60	28.20%	8.87	2.11	23.83%
02:00	23.42	2.92	12.45%	31.03	8.90	28.67%	12.45	3.10	24.88%	9.54	3.09	32.37%
03:00	22.73	2.56	11.25%	36.86	9.43	25.59%	11.58	3.17	27.43%	10.99	2.37	21.57%
04:00	22.96	2.86	12.46%	35.72	11.45	32.05%	13.85	5.95	42.93%	10.56	3.34	31.64%
05:00	23.68	2.76	11.65%	33.11	8.20	24.76%	16.66	9.38	56.32%	11.94	3.31	27.74%
06:00	24.44	2.29	9.38%	28.50	7.27	25.52%	15.60	2.27	14.54%	14.26	3.51	24.64%
07:00	25.26	2.13	8.42%	29.62	6.09	20.56%	17.18	3.17	18.46%	15.66	3.59	22.92%
08:00	26.50	2.20	8.31%	32.93	5.09	15.45%	17.69	2.55	14.39%	18.21	4.14	22.75%
09:00	26.58	2.51	9.46%	33.31	5.87	17.61%	17.20	3.01	17.49%	18.51	2.87	15.52%
10:00	26.37	2.64	10.01%	32.84	6.94	21.12%	16.91	3.17	18.74%	18.08	2.89	15.98%
11:00	24.50	2.78	11.36%	35.28	8.04	22.79%	17.21	2.32	13.49%	18.05	2.65	14.70%
12:00	26.01	3.53	13.56%	33.43	5.27	15.76%	17.49	2.92	16.70%	18.05	2.88	15.93%
13:00	26.04	3.03	11.65%	33.60	5.31	15.81%	18.10	1.54	8.50%	18.91	3.36	17.76%
14:00	25.64	2.24	8.74%	33.39	3.76	11.26%	18.44	2.42	13.15%	18.25	2.93	16.06%
15:00	24.28	2.24	9.22%	31.61	7.66	24.25%	17.10	2.49	14.55%	17.08	1.82	10.66%
16:00	23.37	2.14	9.15%	33.62	10.12	30.11%	14.44	3.30	22.84%	15.96	1.83	11.44%
17:00	24.89	2.67	10.72%	36.36	11.16	30.69%	16.08	3.05	18.99%	16.04	2.31	14.42%
18:00	25.03	2.49	9.96%	42.88	12.95	30.20%	15.70	3.37	21.48%	15.54	3.08	19.83%
19:00	24.90	2.09	8.38%	37.89	7.70	20.31%	15.52	1.87	12.07%	14.47	2.85	19.67%
20:00	24.18	2.33	9.64%	39.89	8.68	21.75%	15.85	6.16	38.89%	14.10	2.66	18.87%
21:00	24.35	2.32	9.52%	40.73	9.49	23.29%	13.67	5.43	39.72%	14.19	2.28	16.10%
22:00	23.72	2.36	9.93%	44.63	9.09	20.38%	13.22	3.84	29.08%	11.90	2.19	18.42%
23:00	23.23	1.95	8.40%	39.70	13.61	34.29%	14.44	5.24	36.31%	13.15	4.22	32.13%
<b>TOTAL</b>	<b>24.65</b>	<b>2.59</b>	<b>10.53%</b>	<b>34.95</b>	<b>8.33</b>	<b>23.75%</b>	<b>15.41</b>	<b>3.65</b>	<b>24.54%</b>	<b>14.68</b>	<b>2.87</b>	<b>20.47%</b>

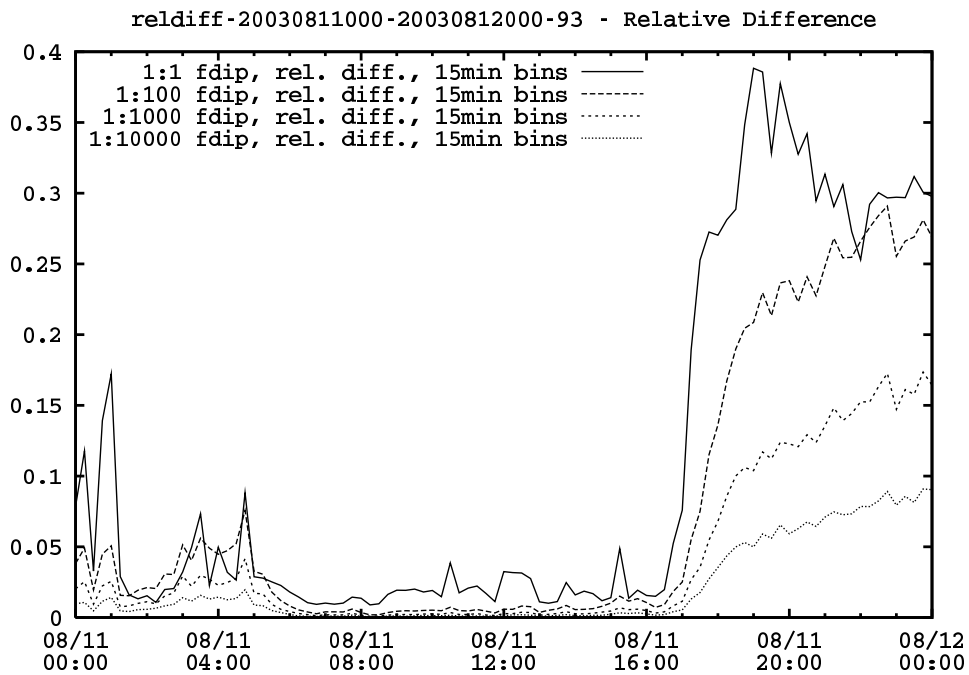
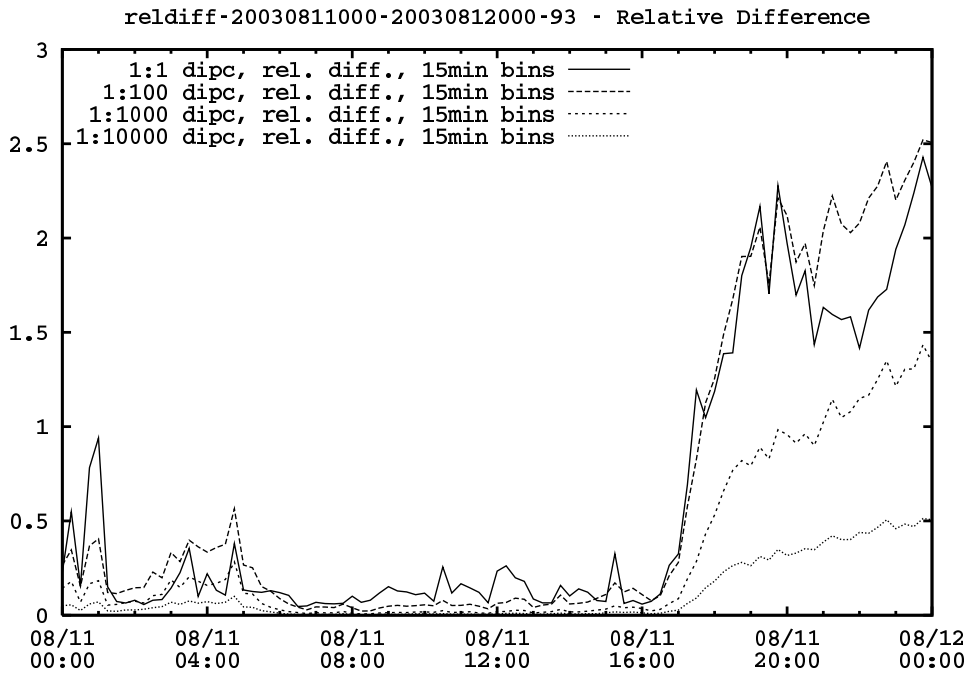


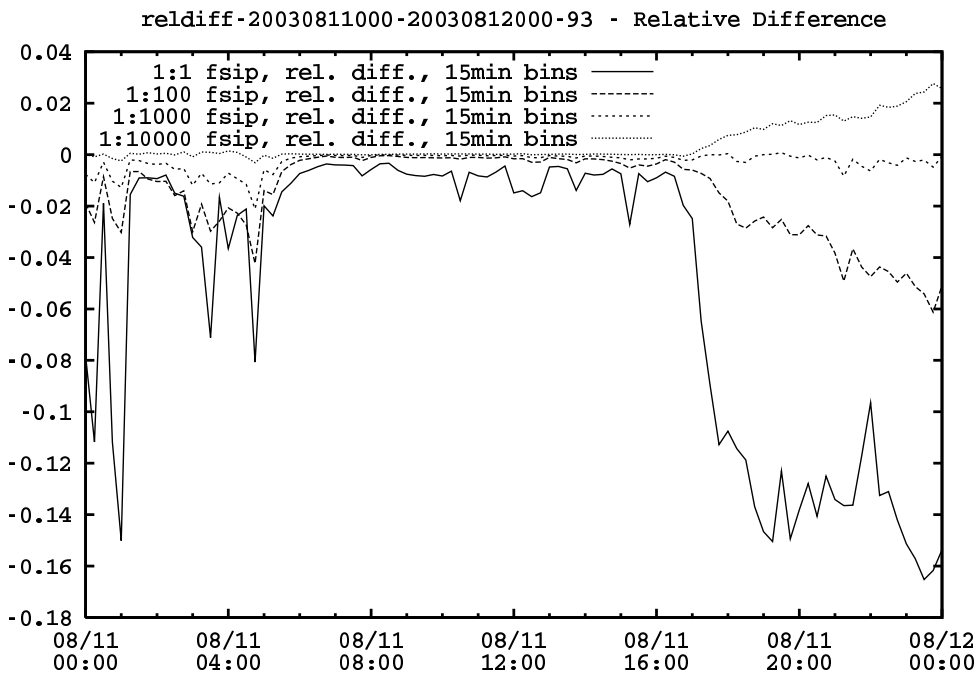
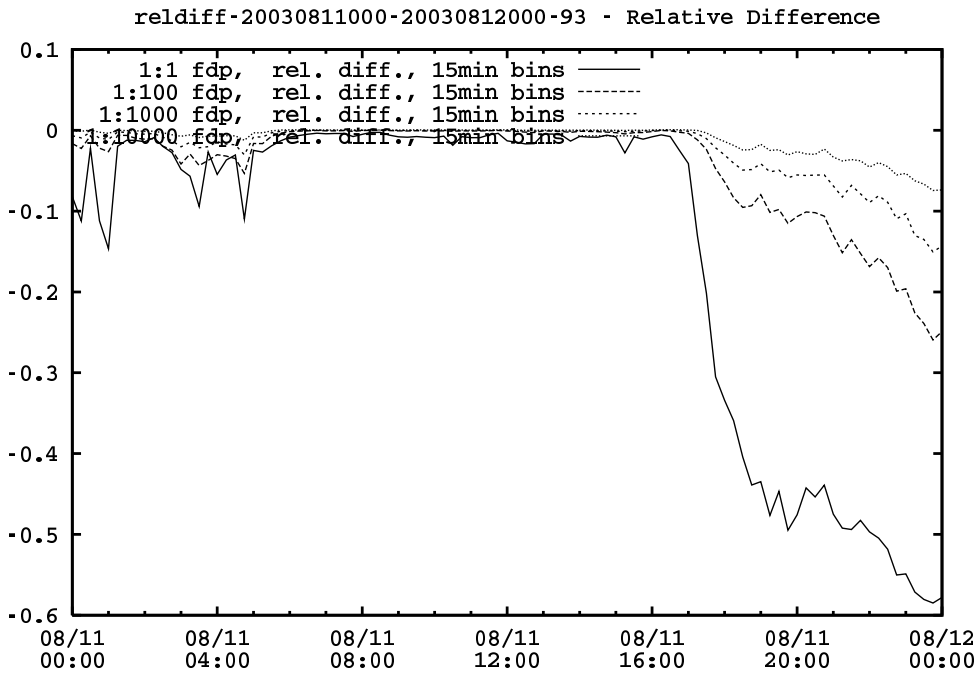


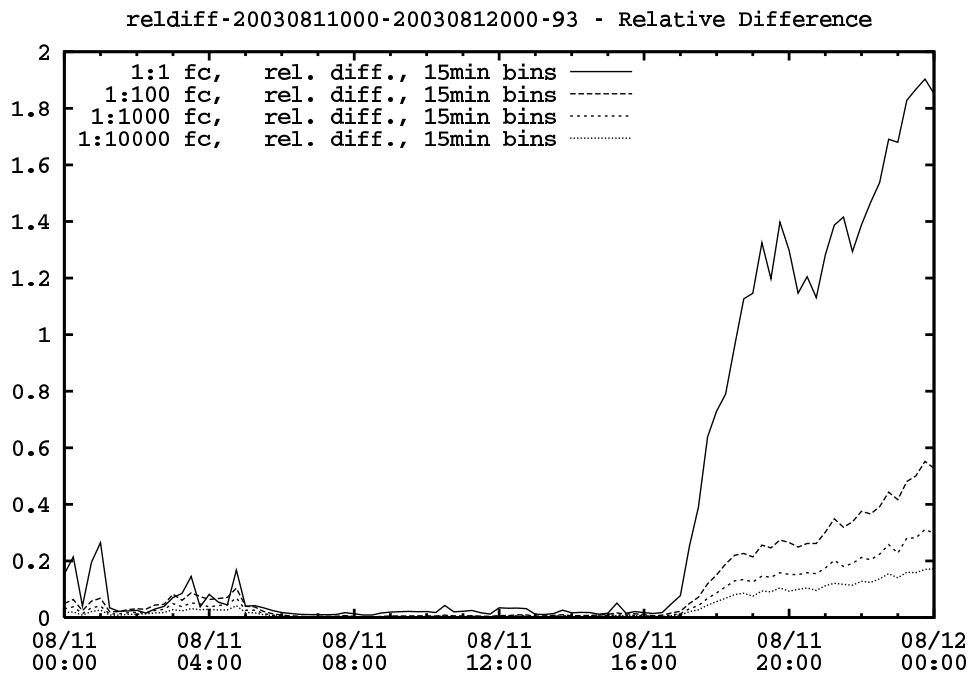
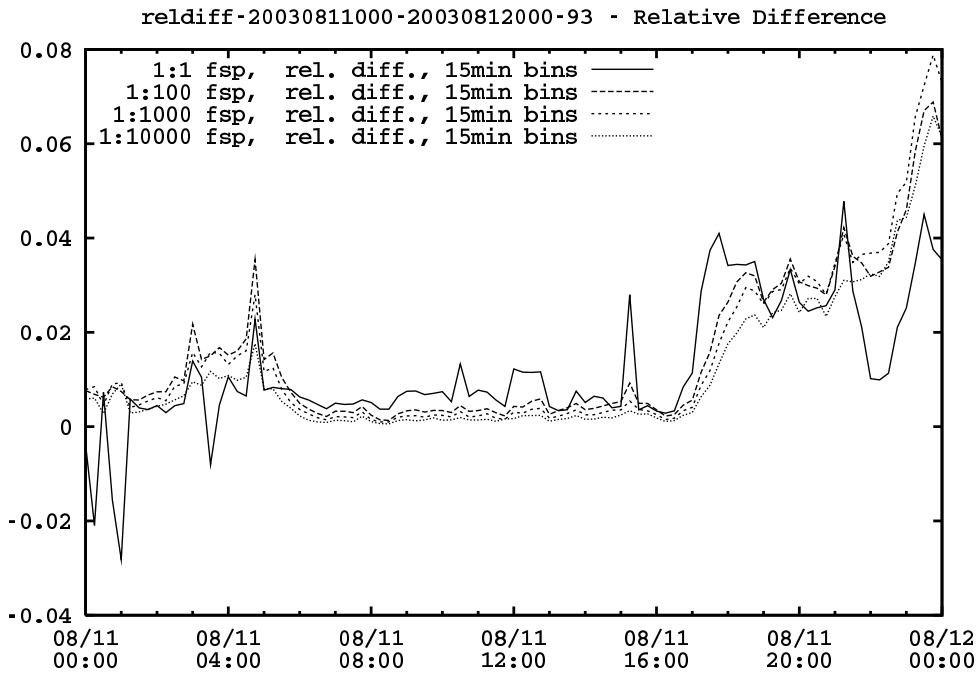
## B Relative difference plots for Blaster

### B.1 Router 1

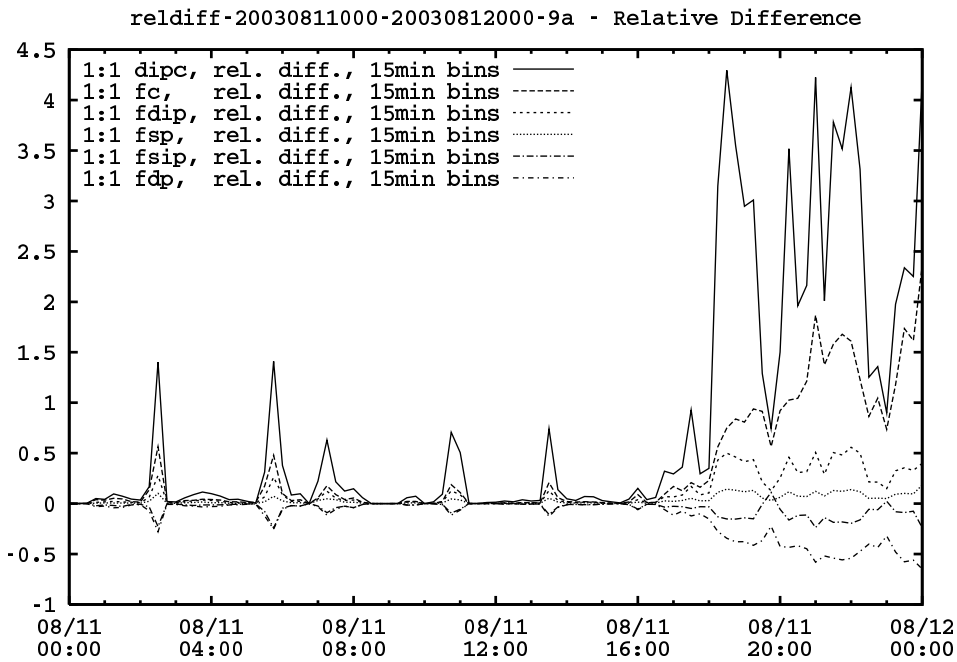
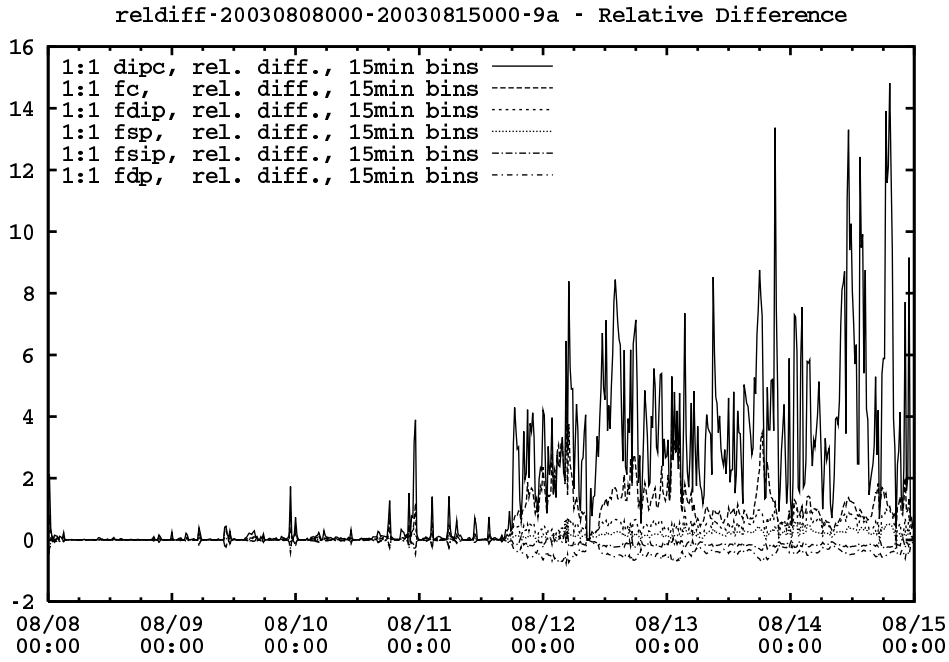


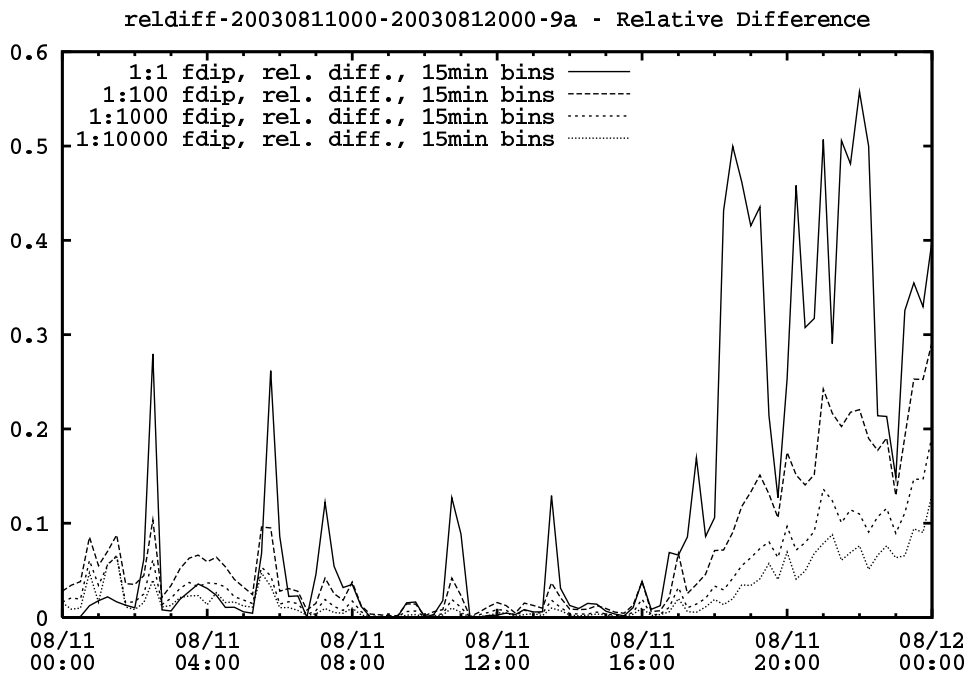
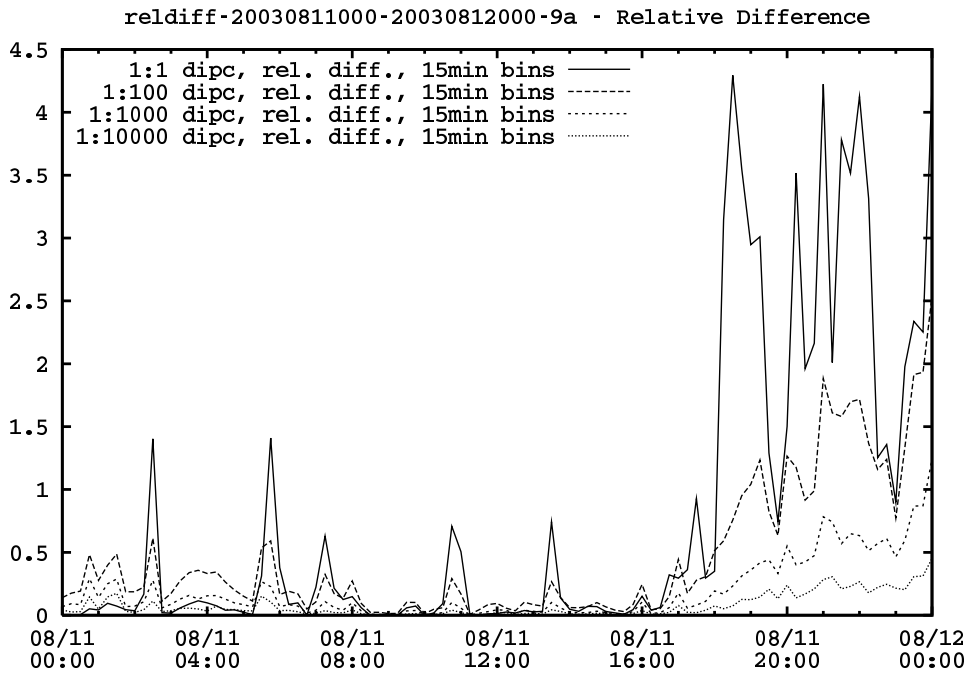


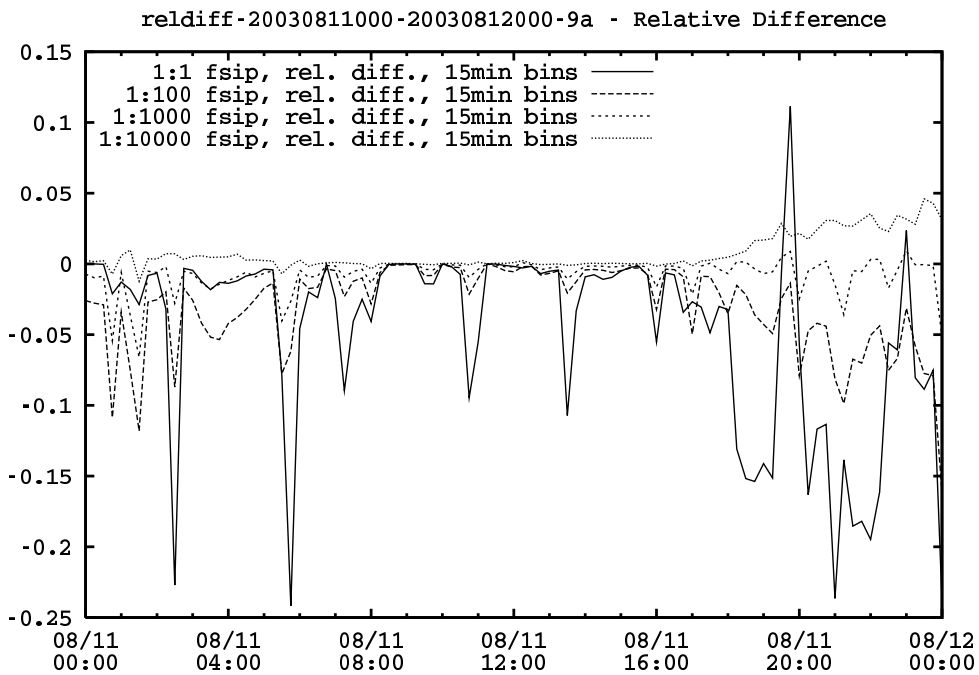
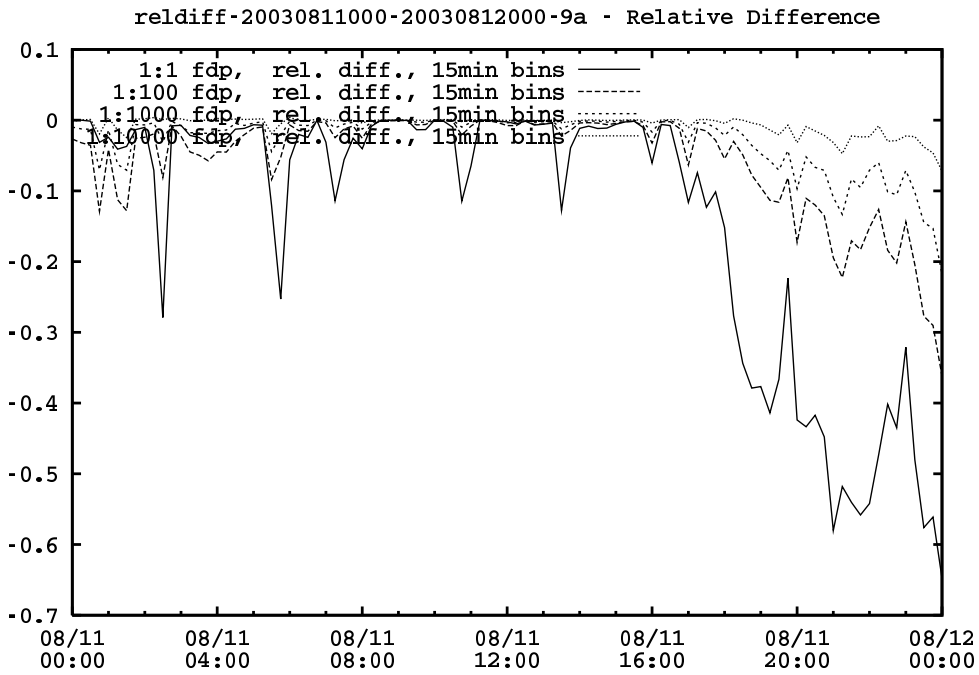




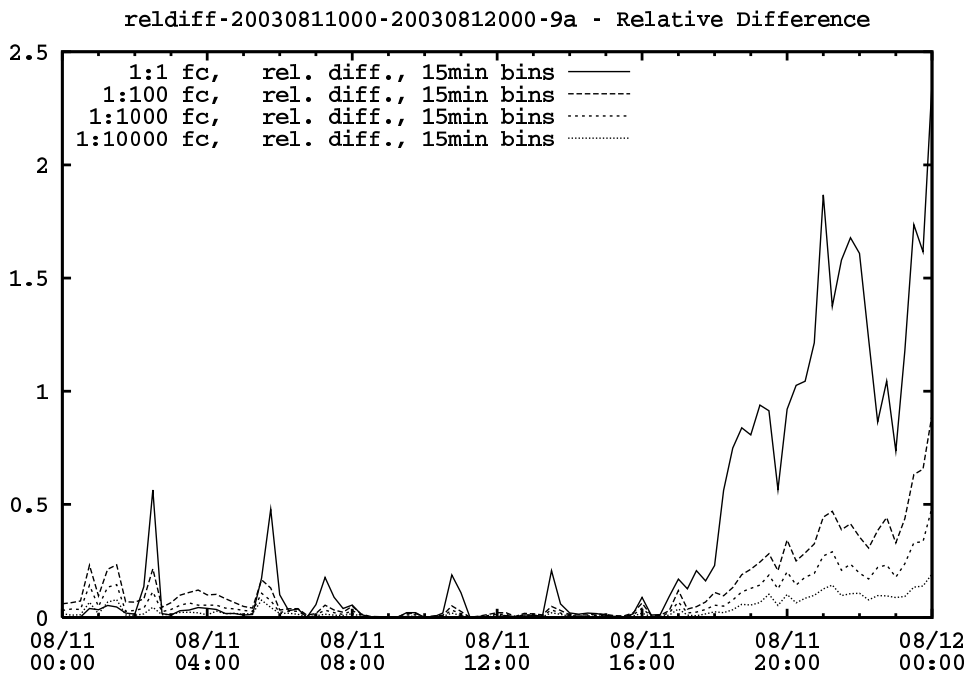
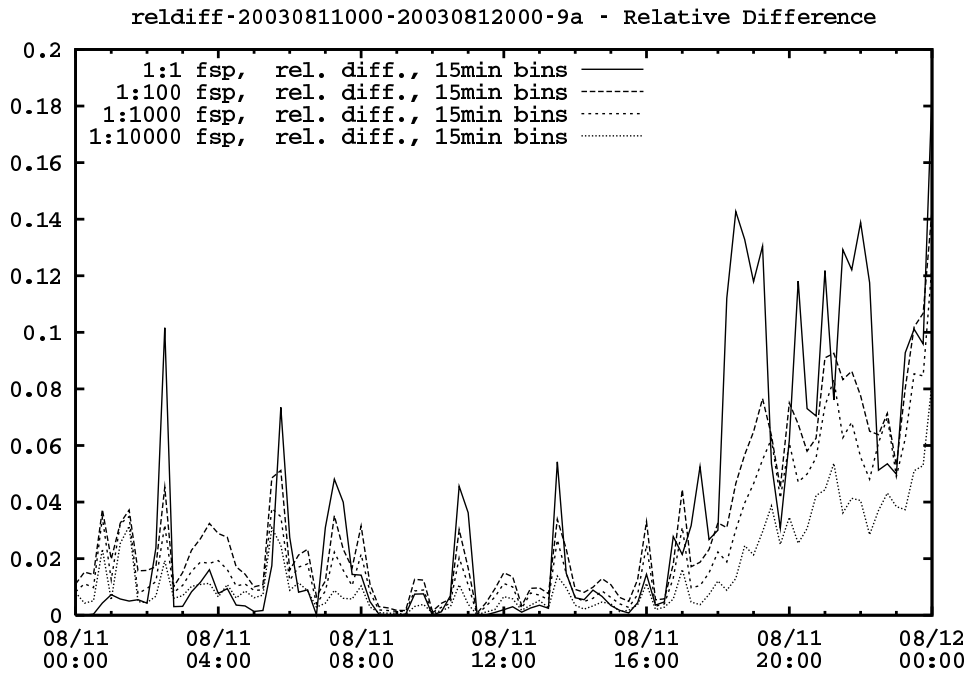
## B.2 Router 2



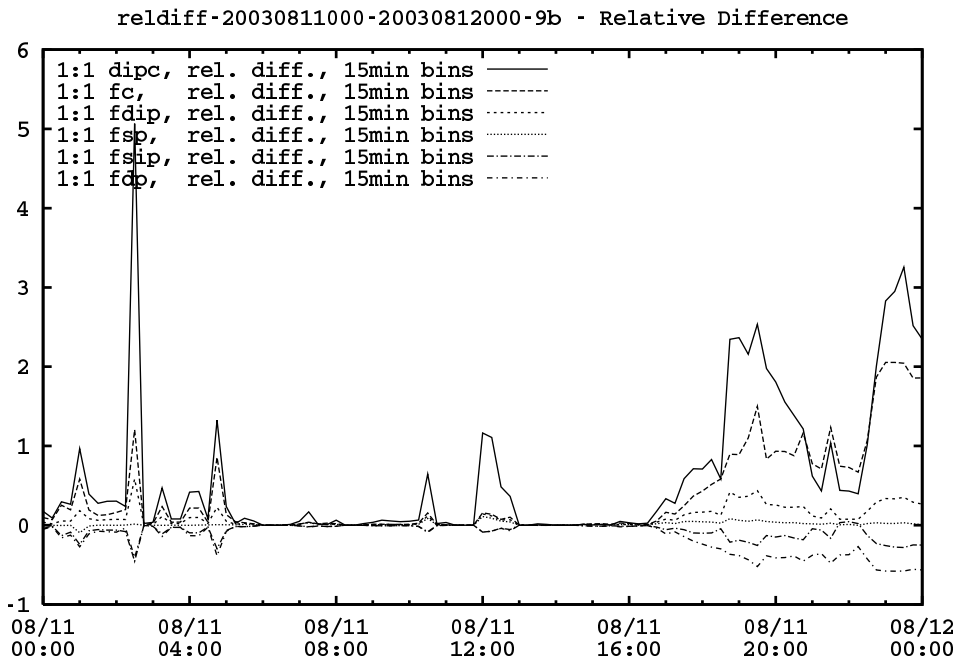
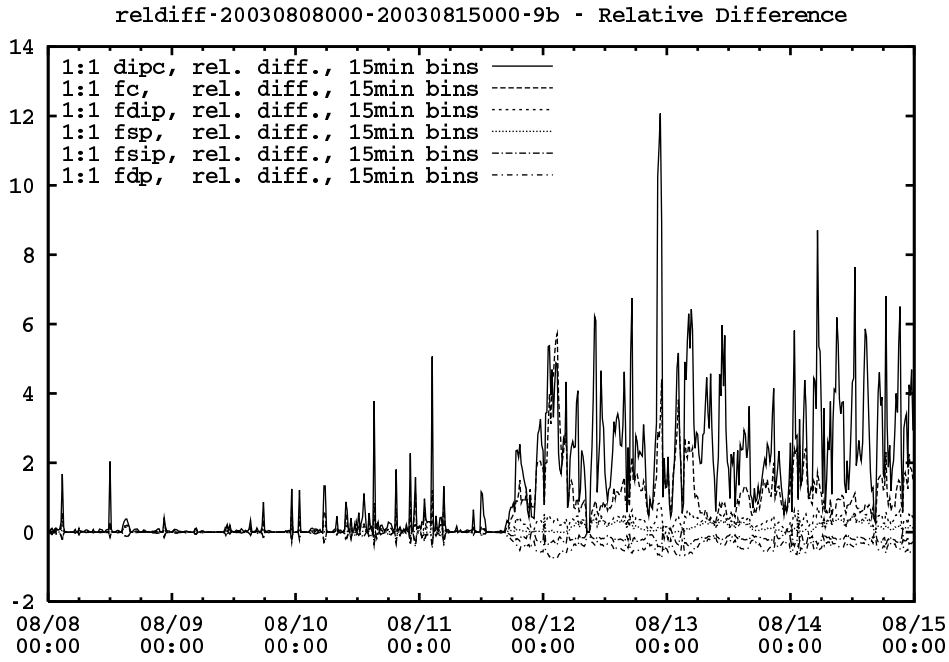


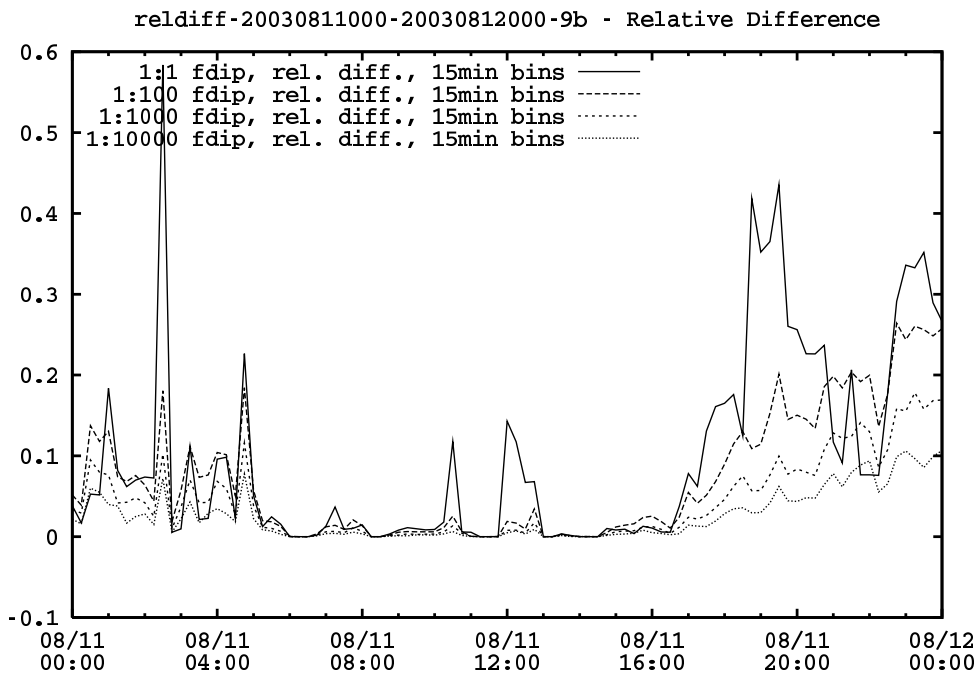
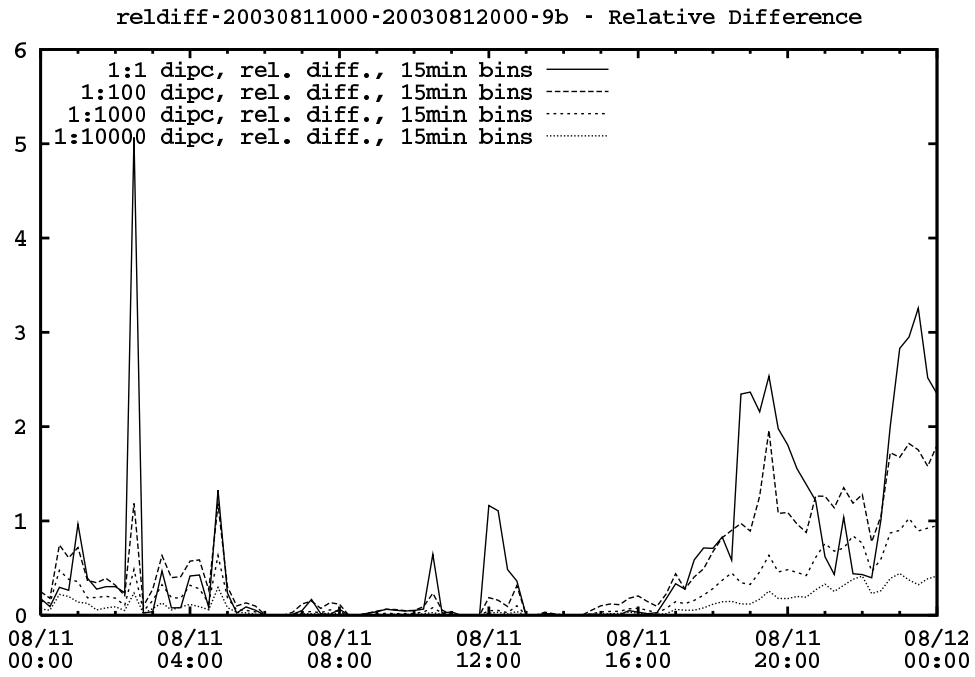


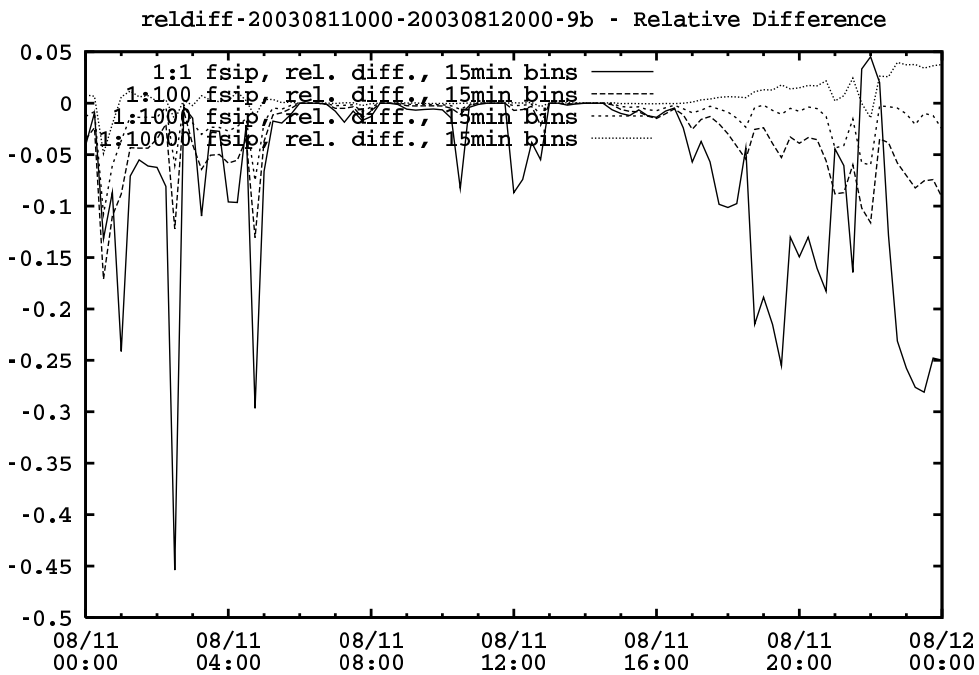
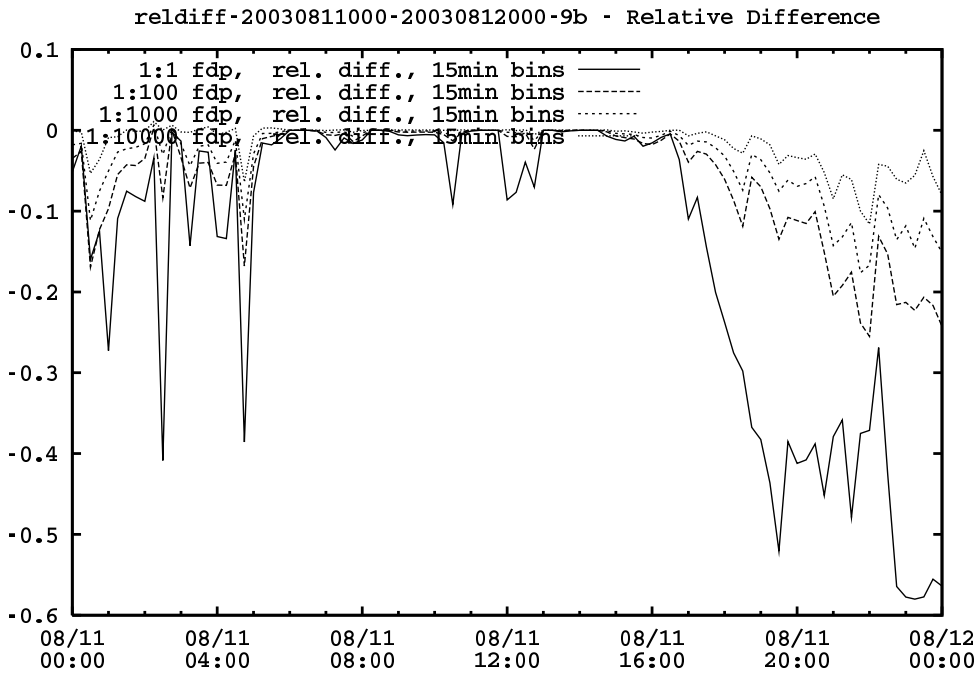


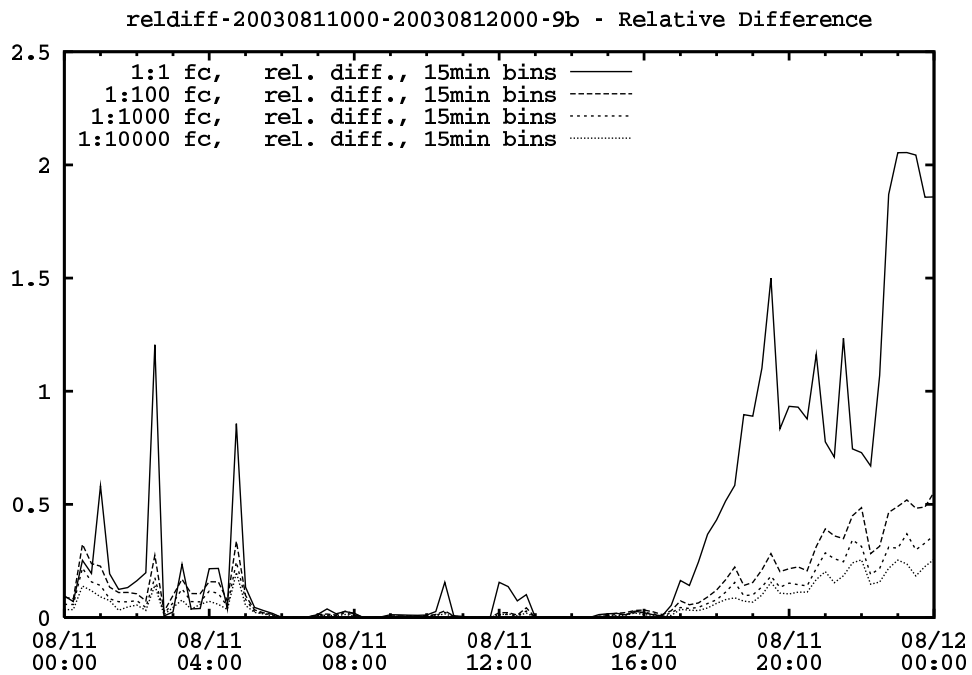
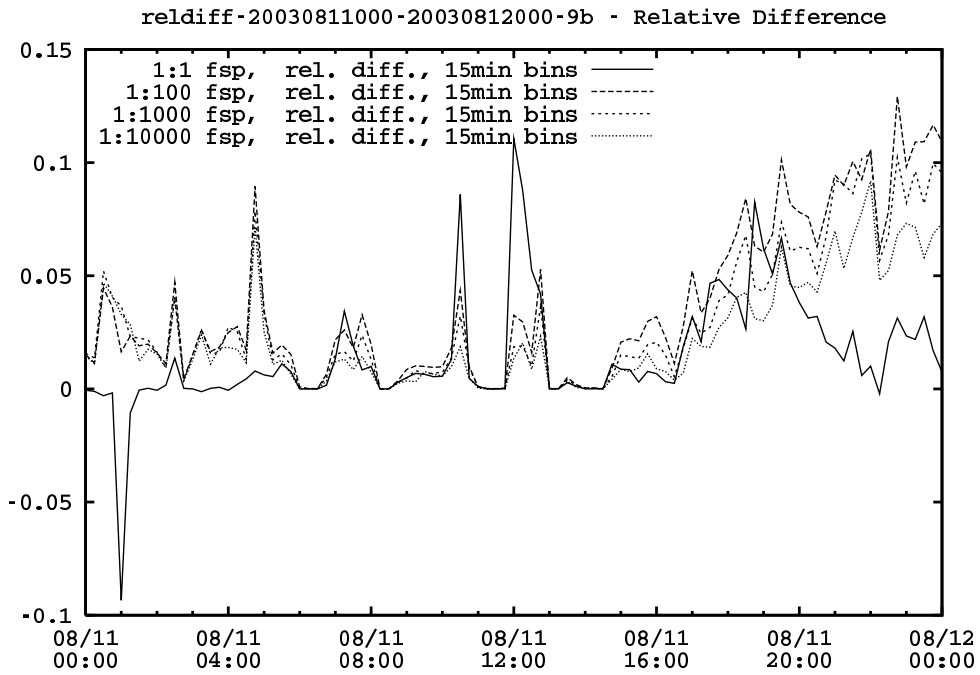


### B.3 Router 3

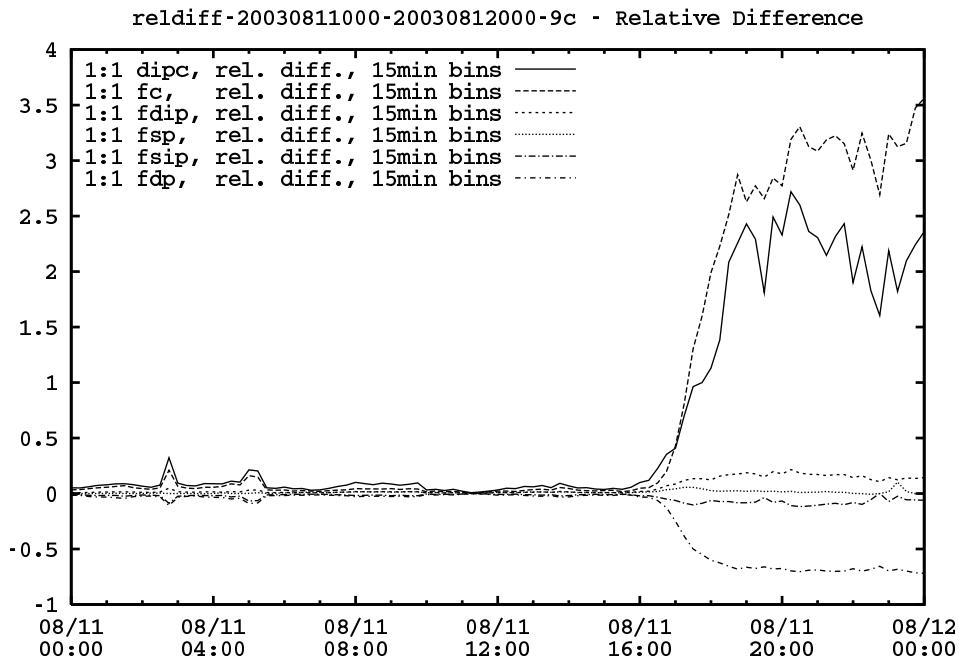
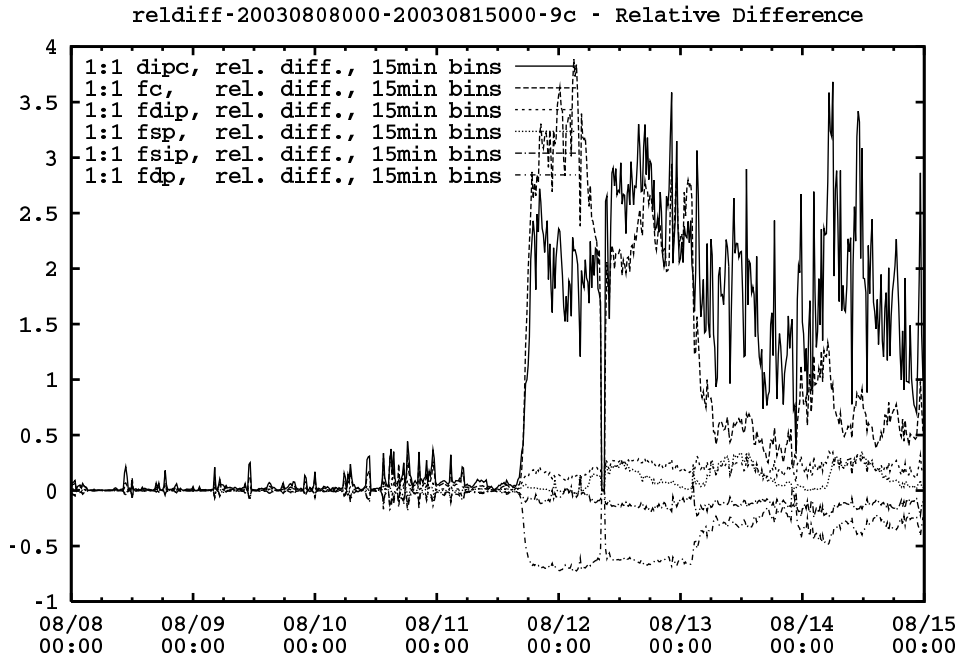


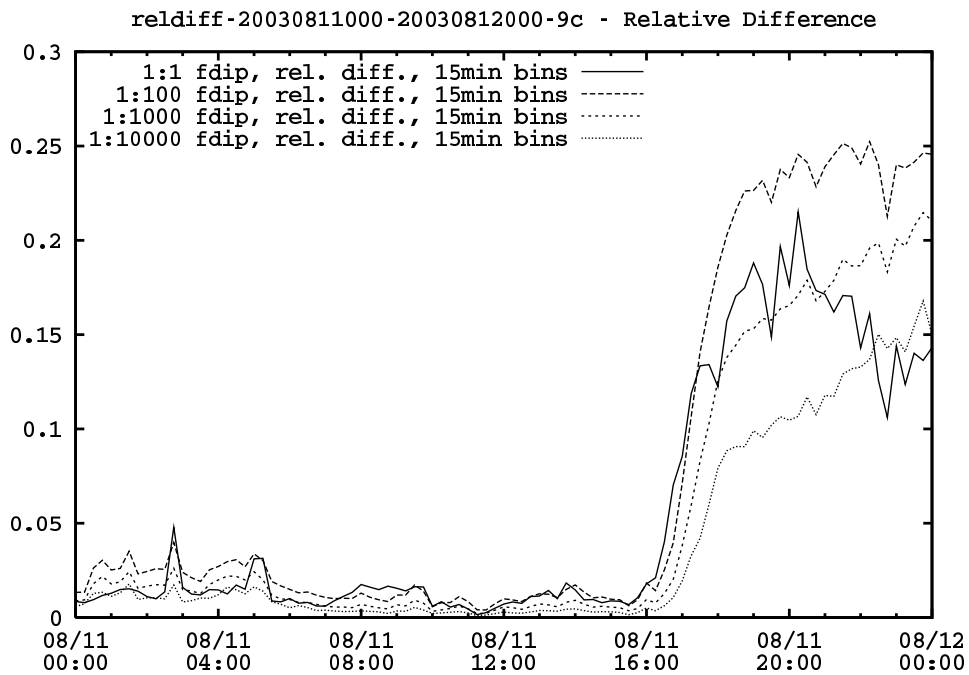
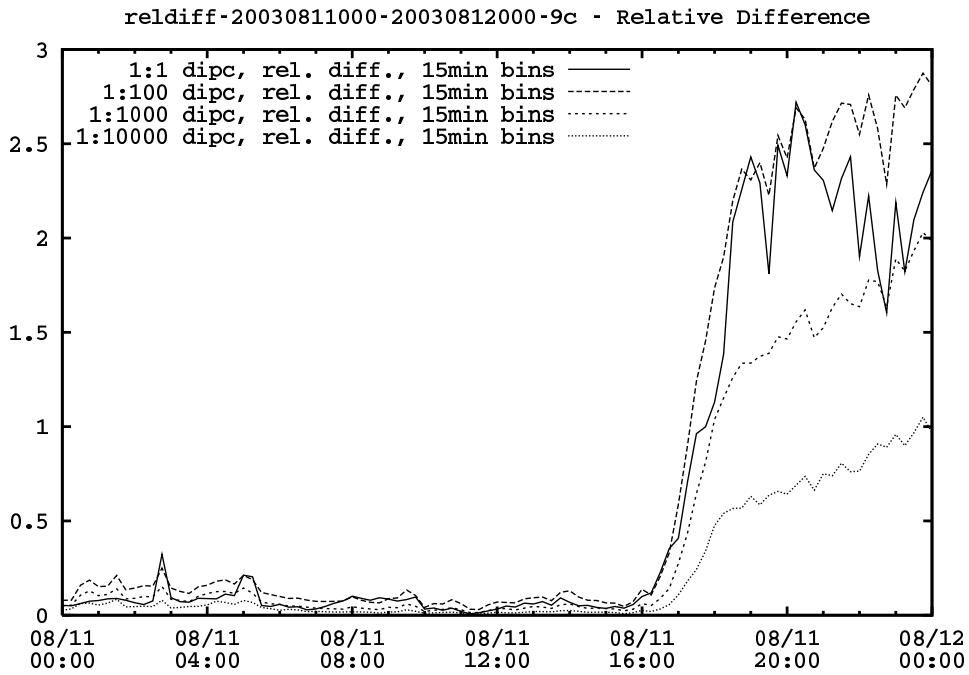


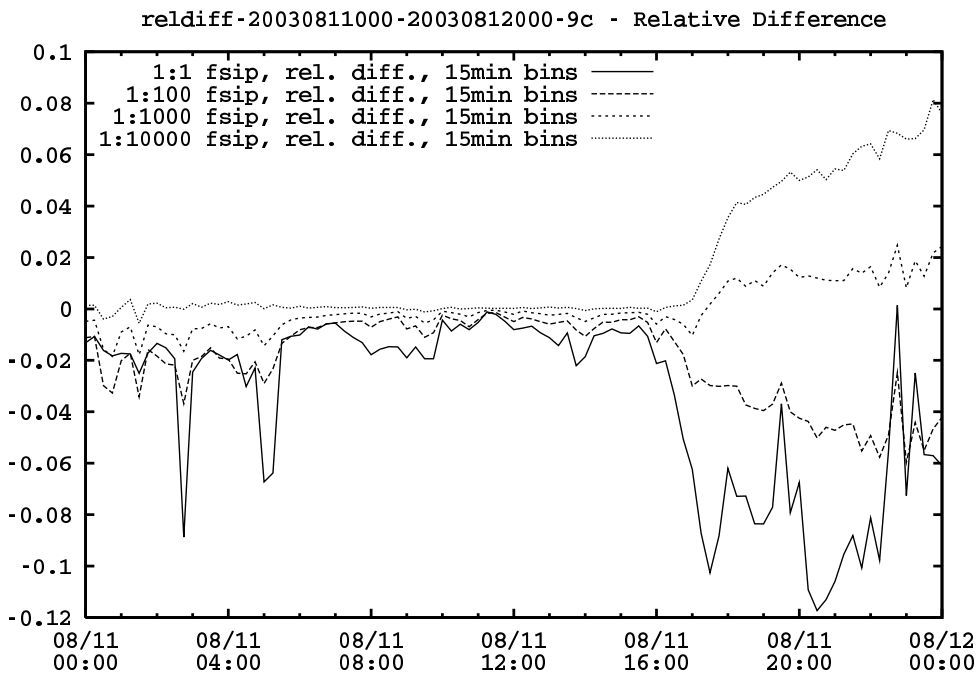
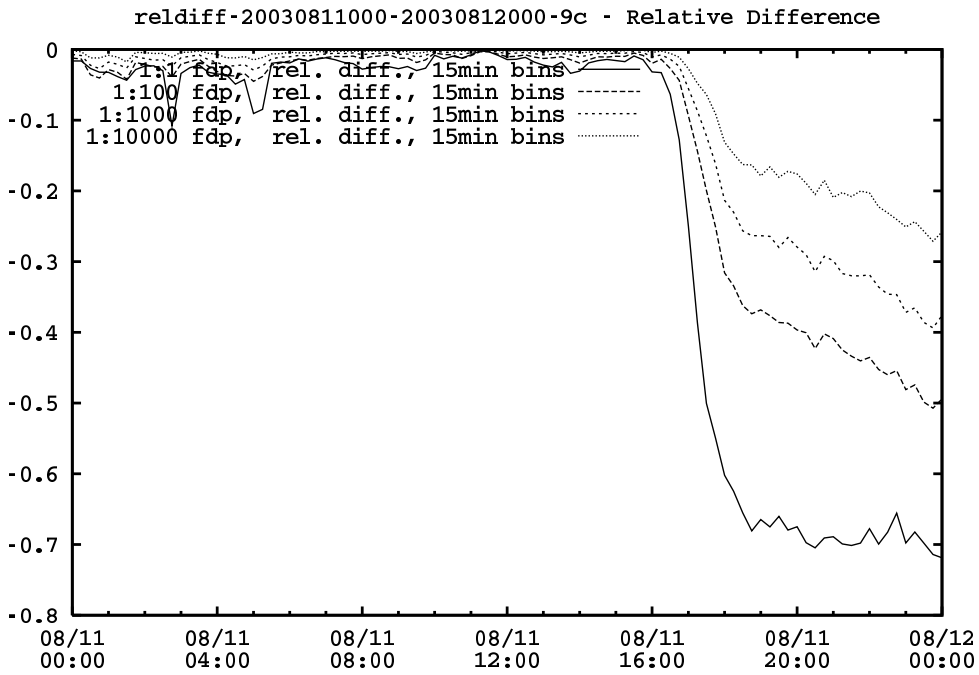




## B.4 Router 4

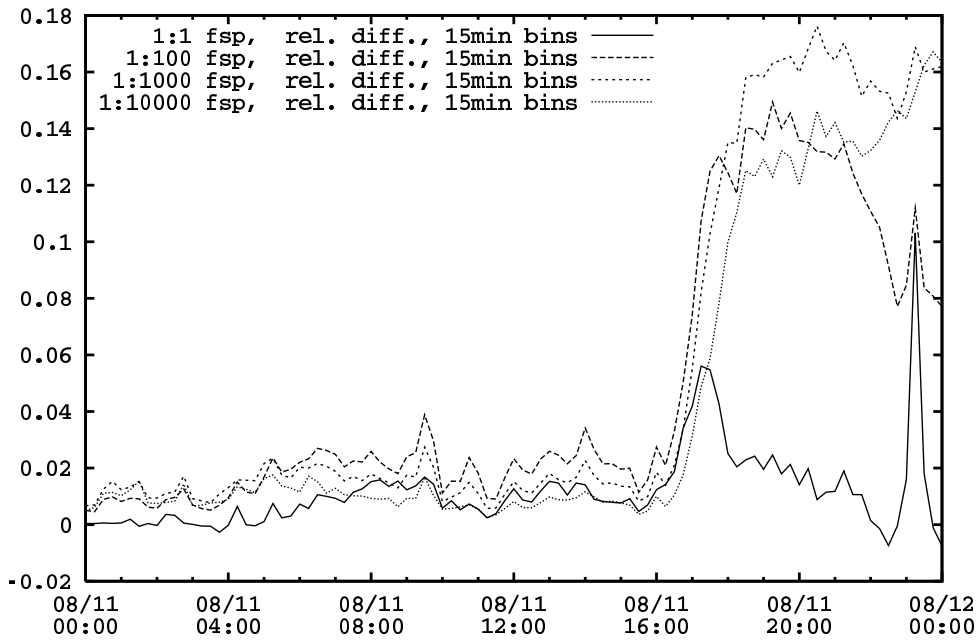






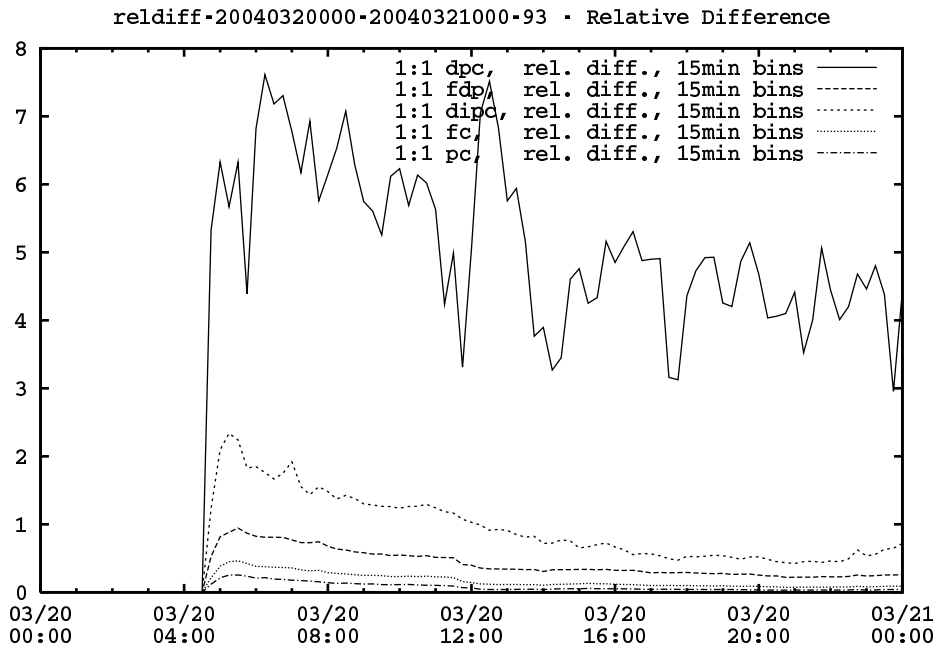
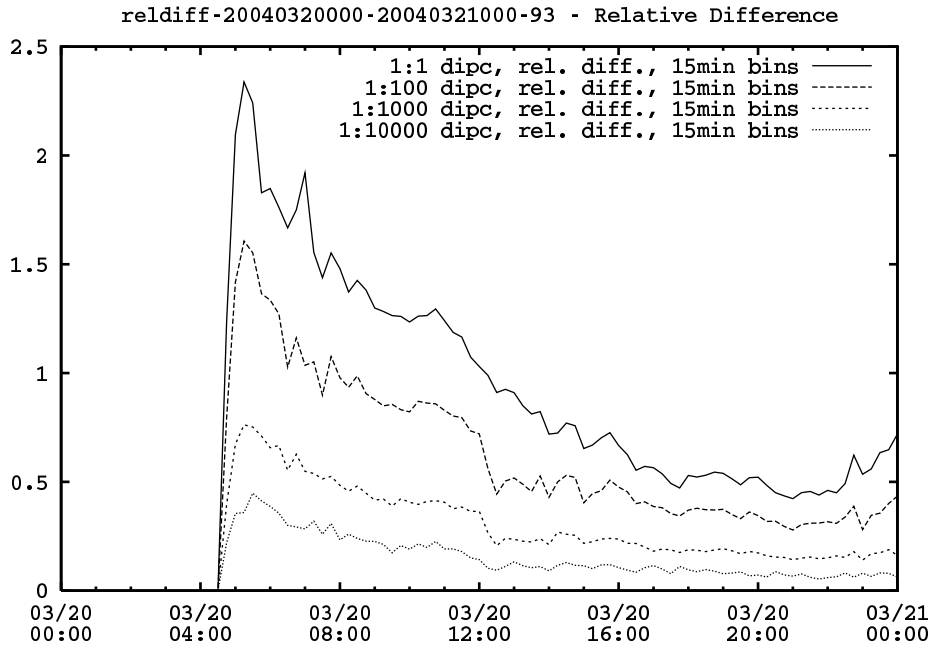


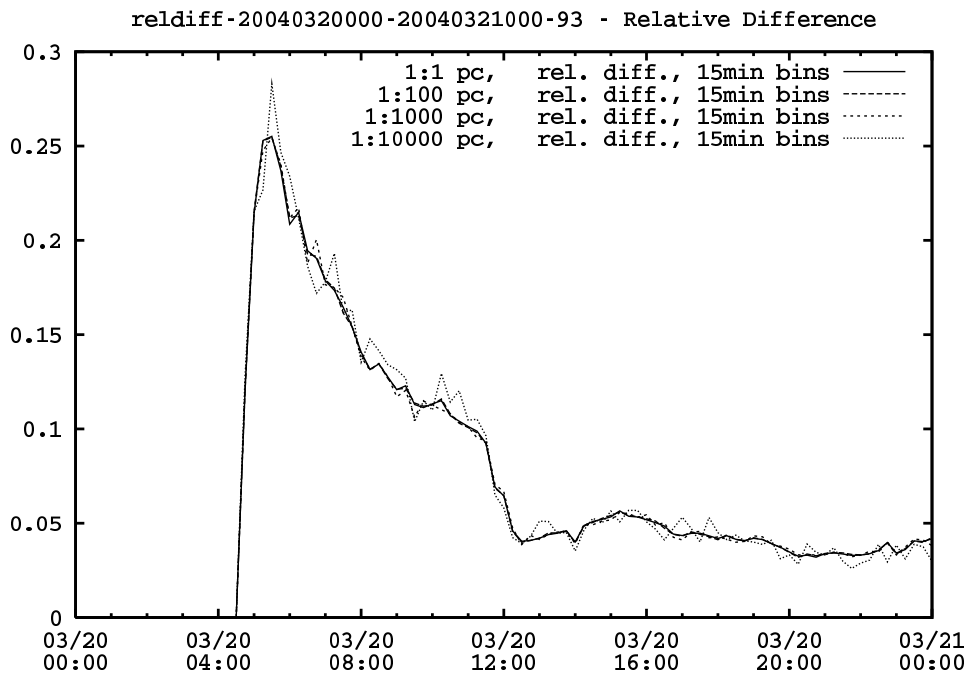
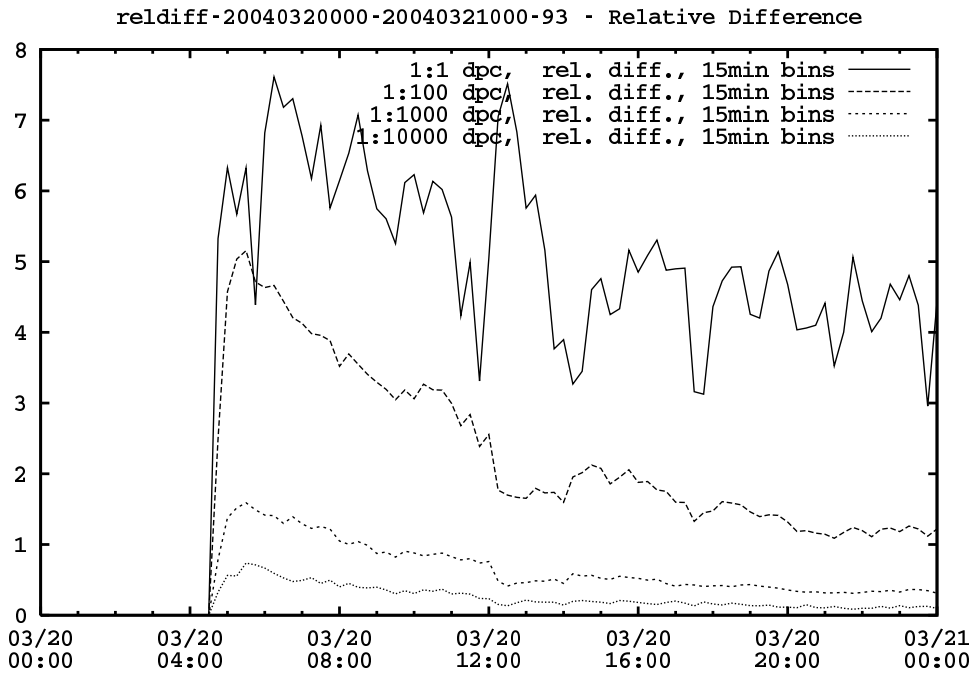
reldiff-20030811000-20030812000-9c - Relative Difference

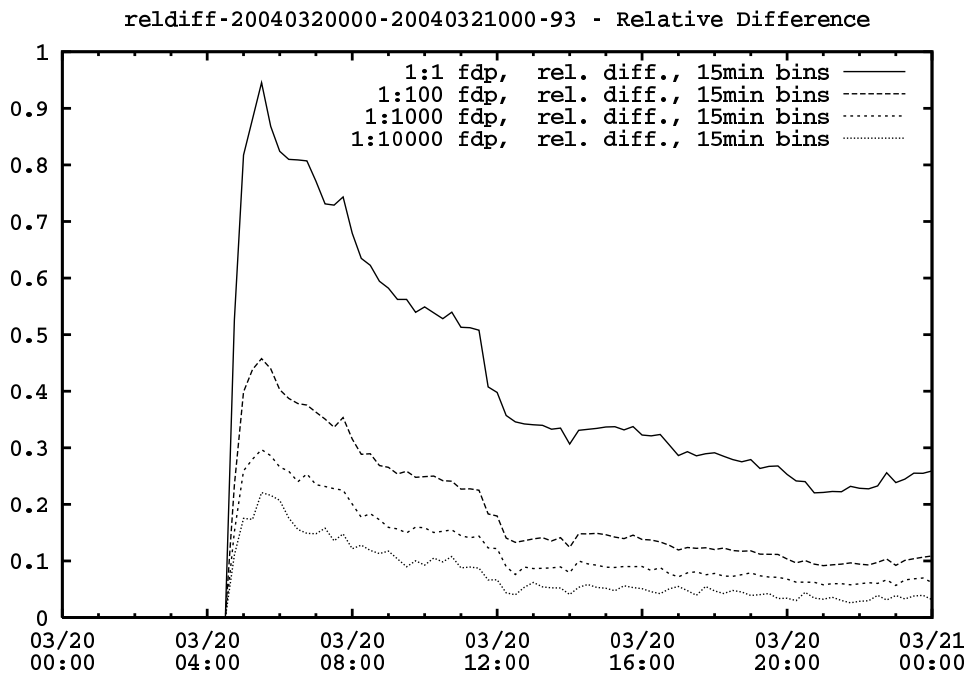
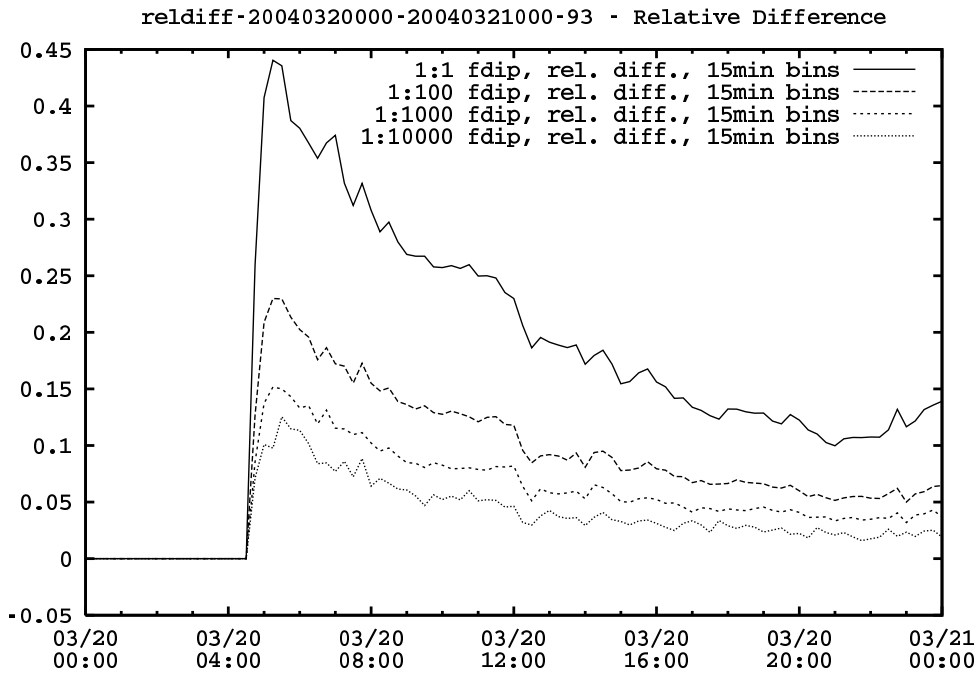


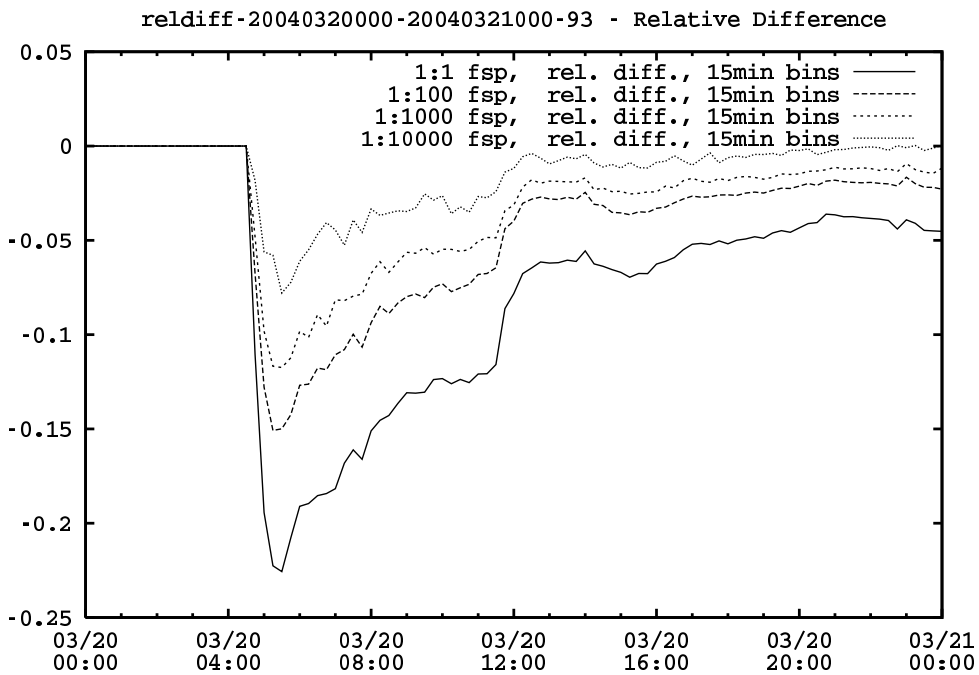
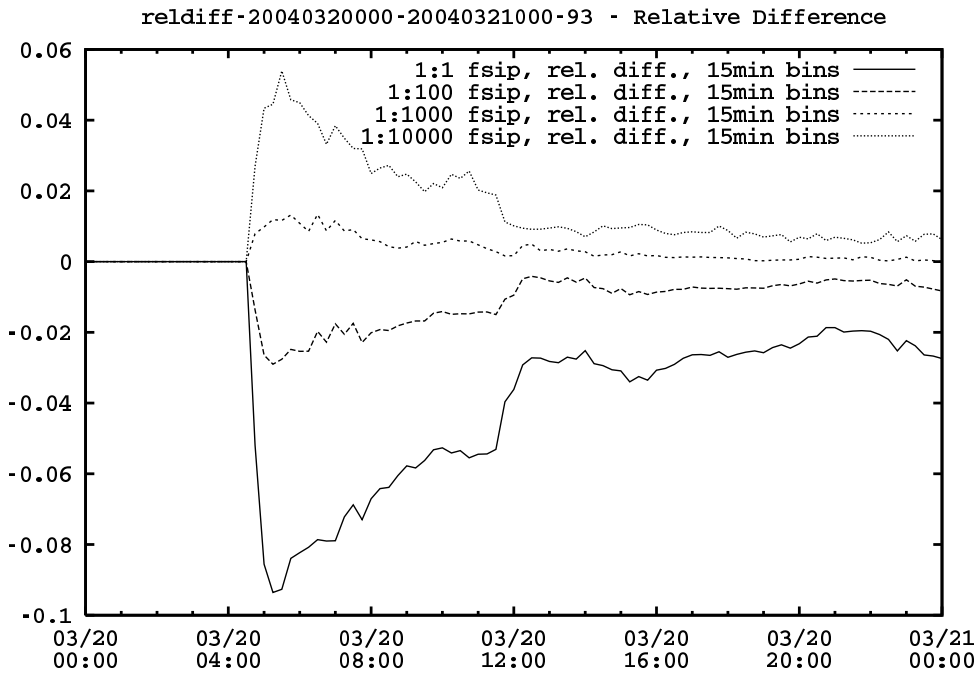
# C Relative difference plots for Witty

## C.1 Router 1

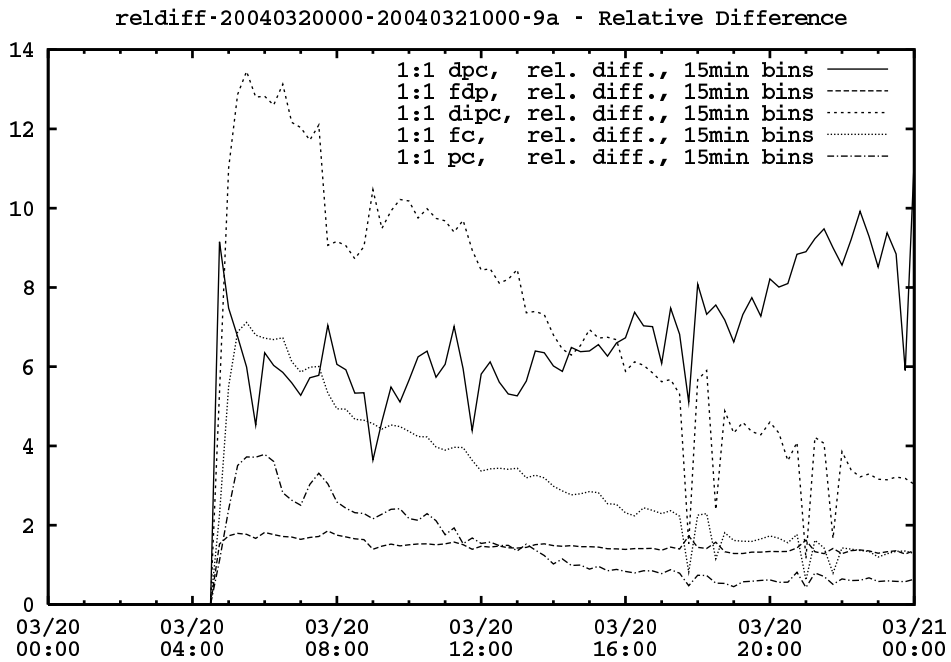
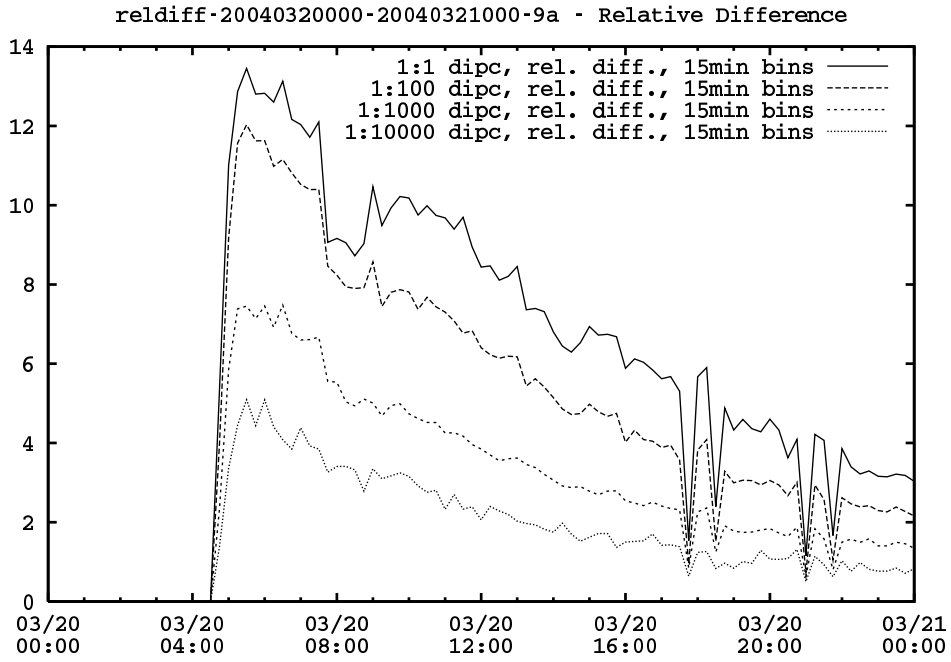




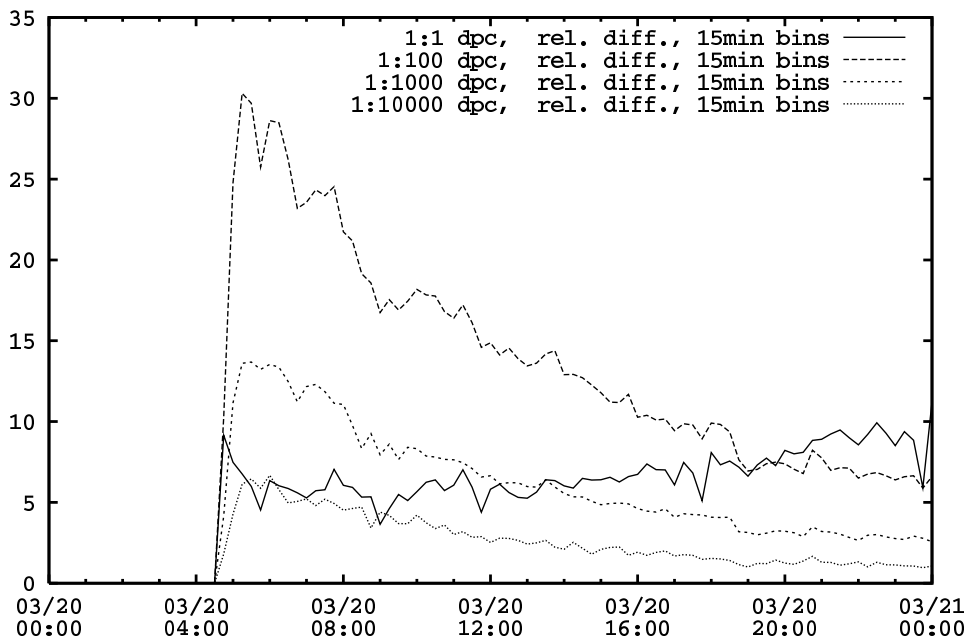




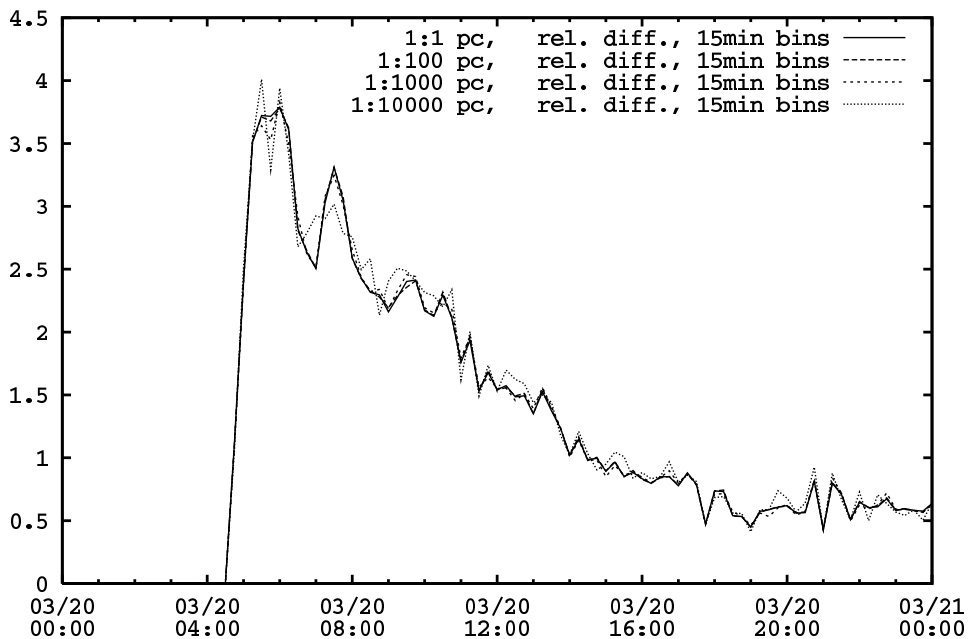
## C.2 Router 2

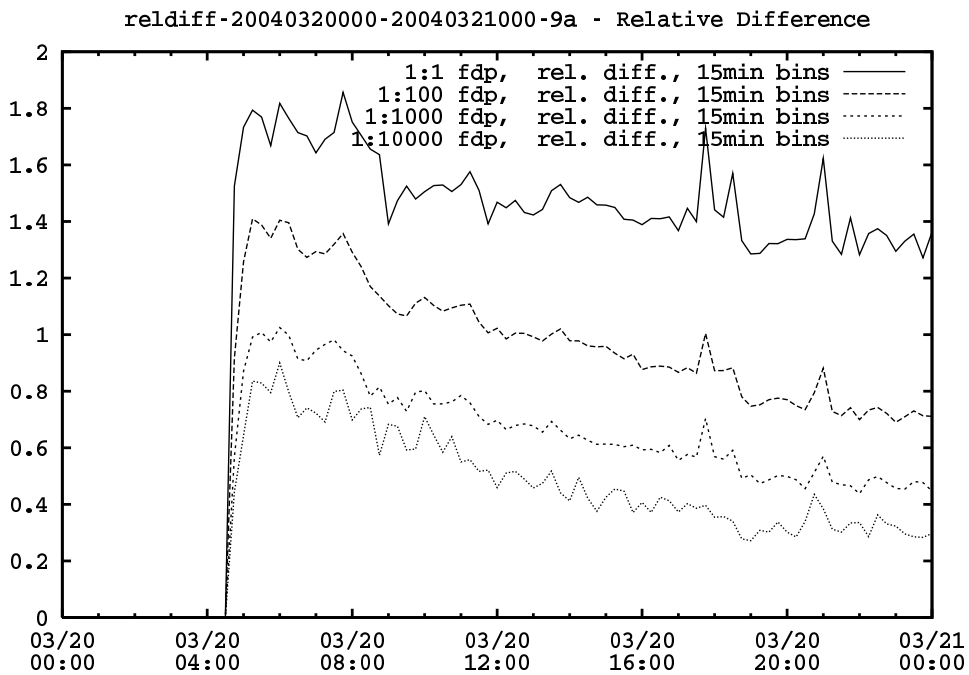
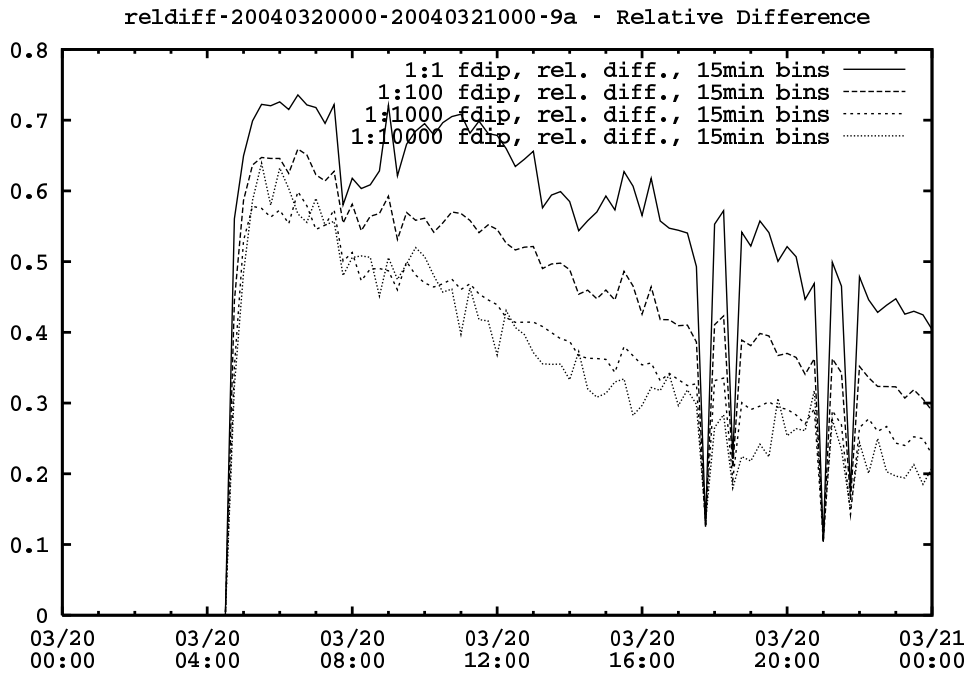


reldiff-20040320000-20040321000-9a - Relative Difference

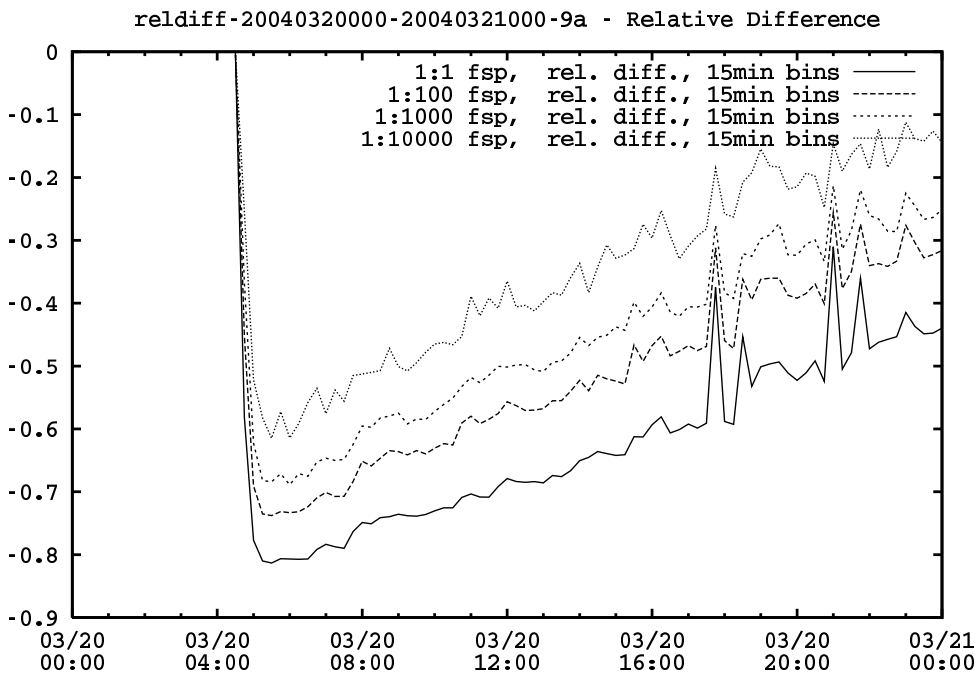
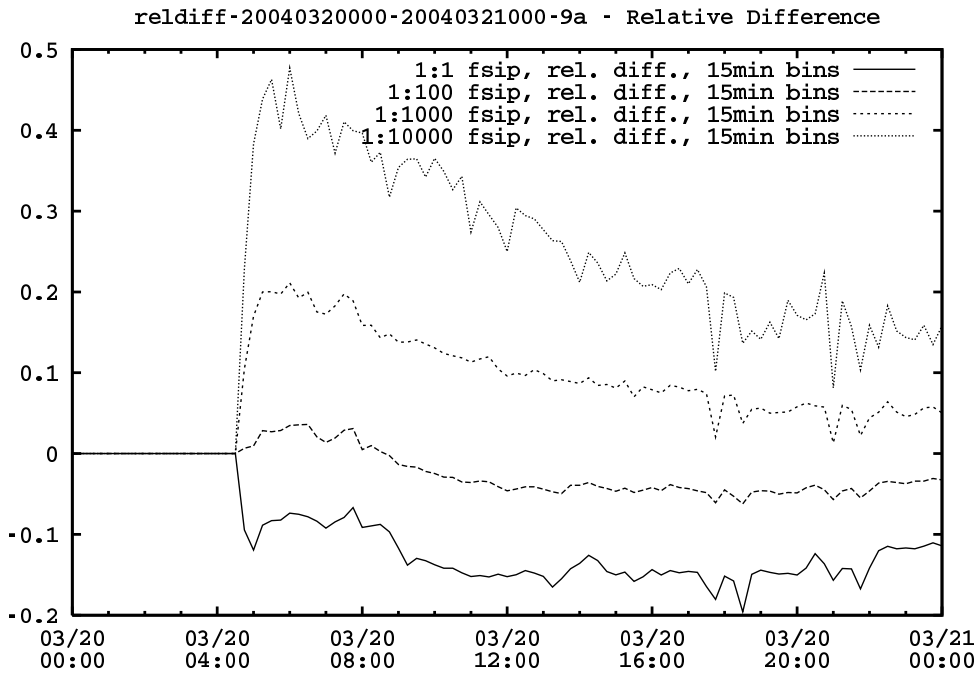


reldiff-20040320000-20040321000-9a - Relative Difference

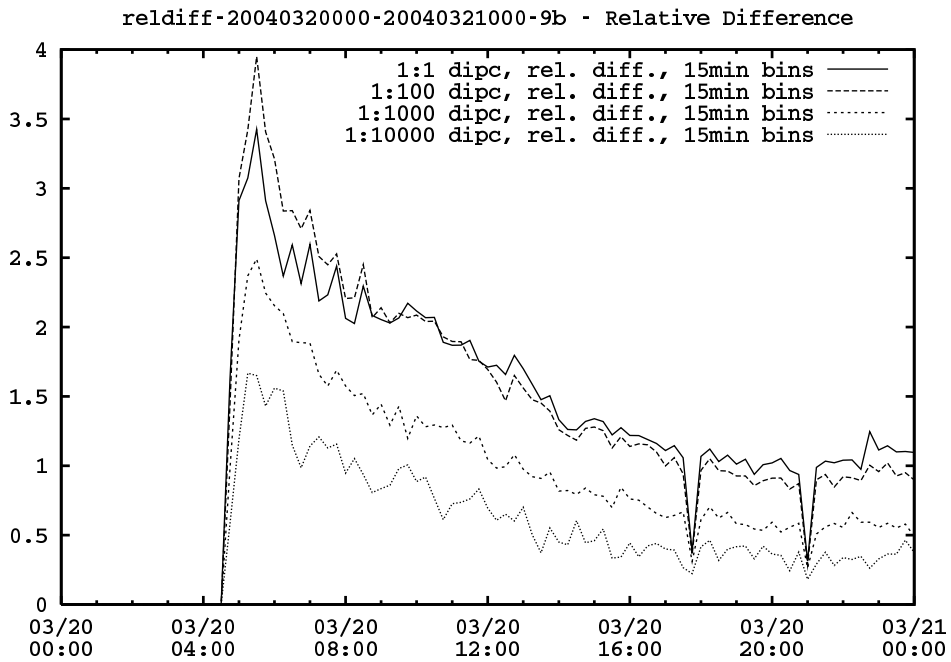
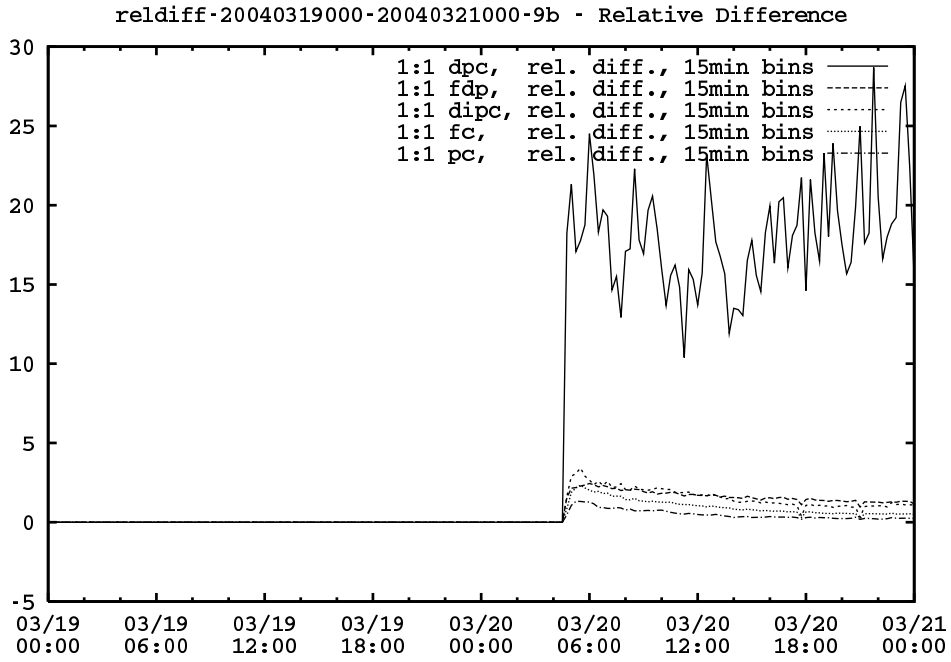


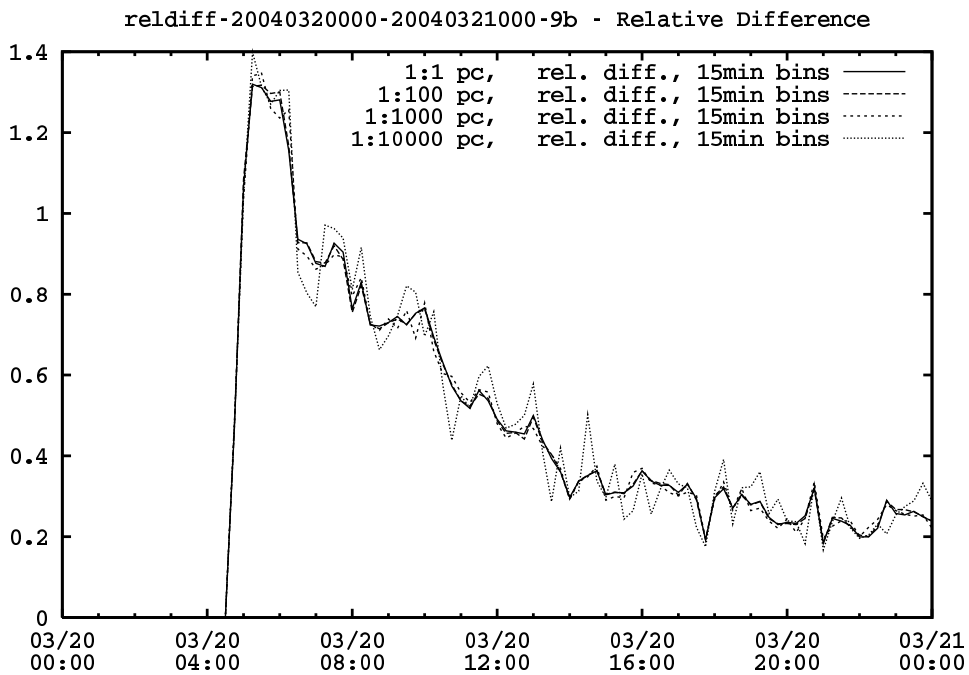
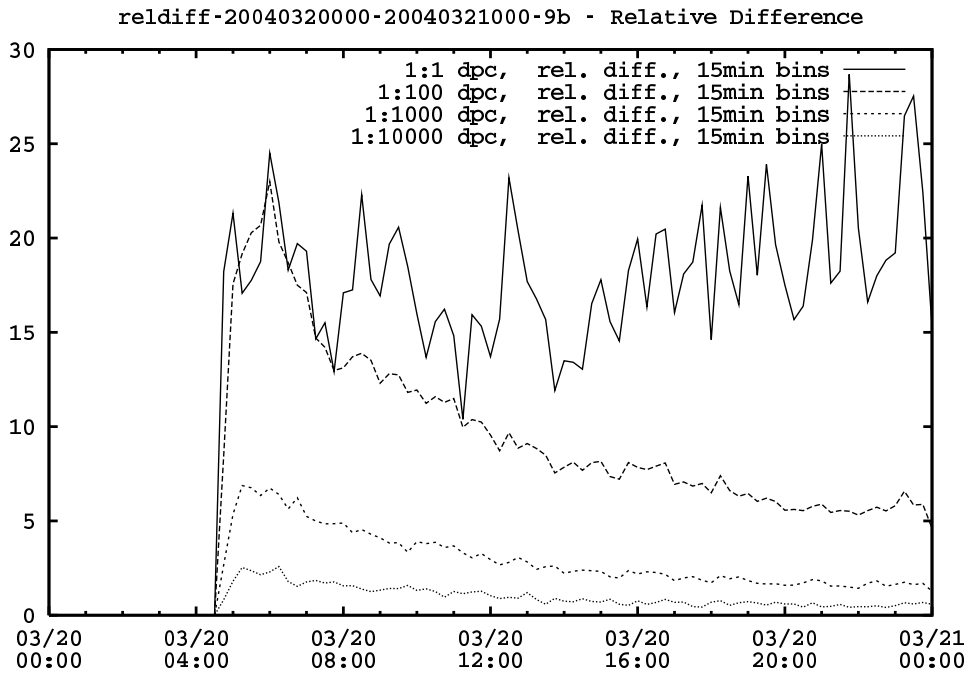


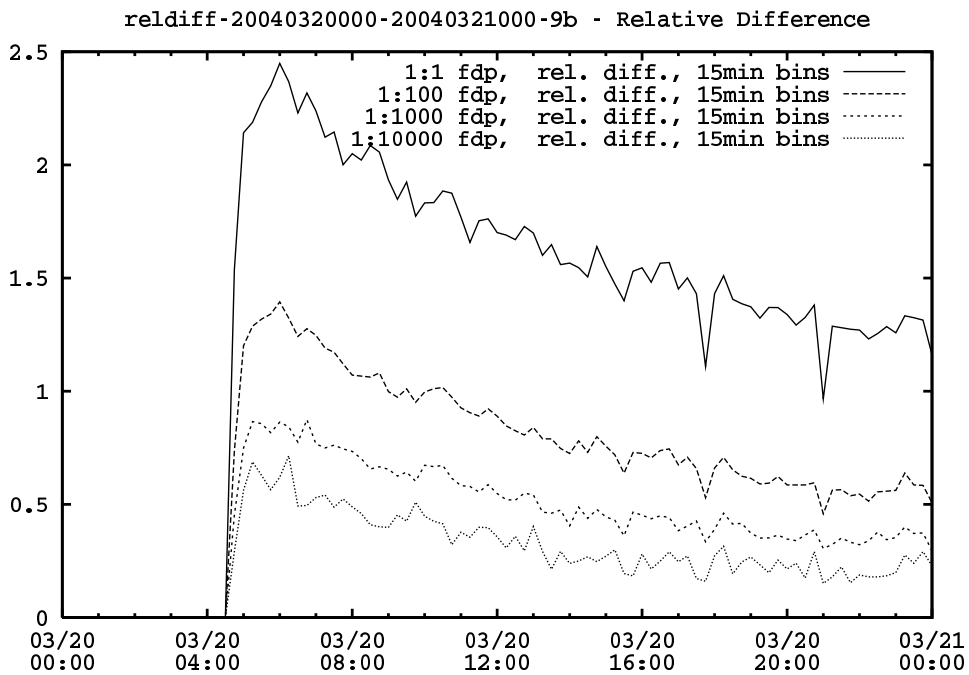
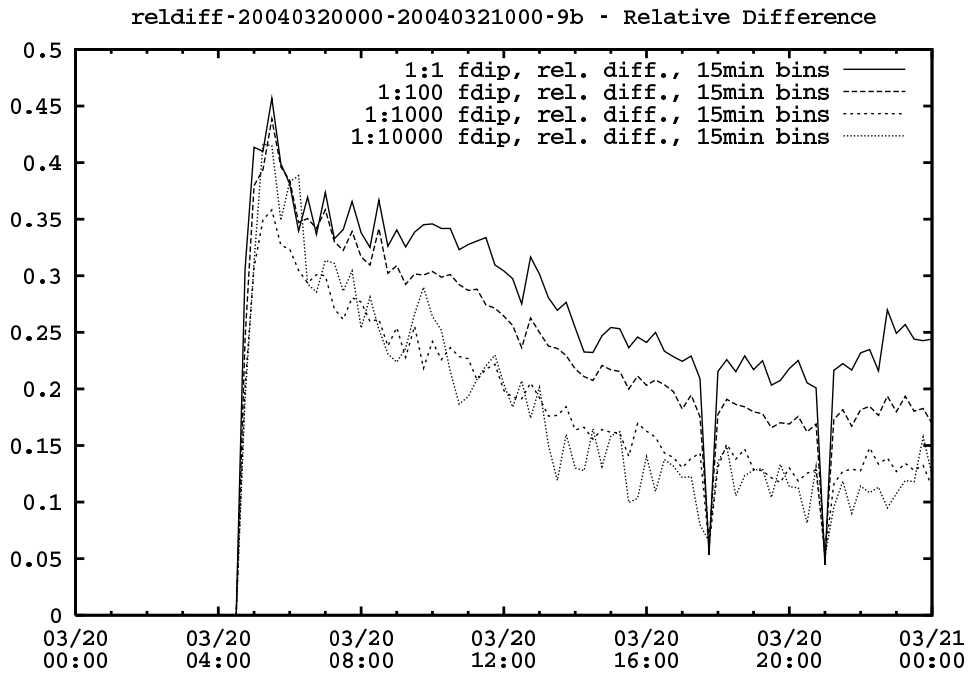




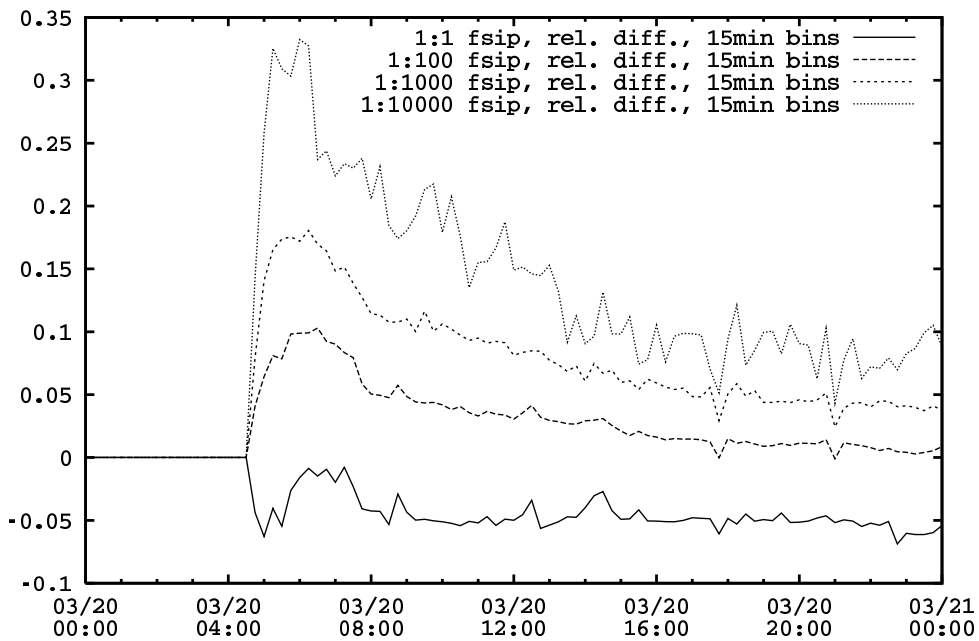
### C.3 Router 3



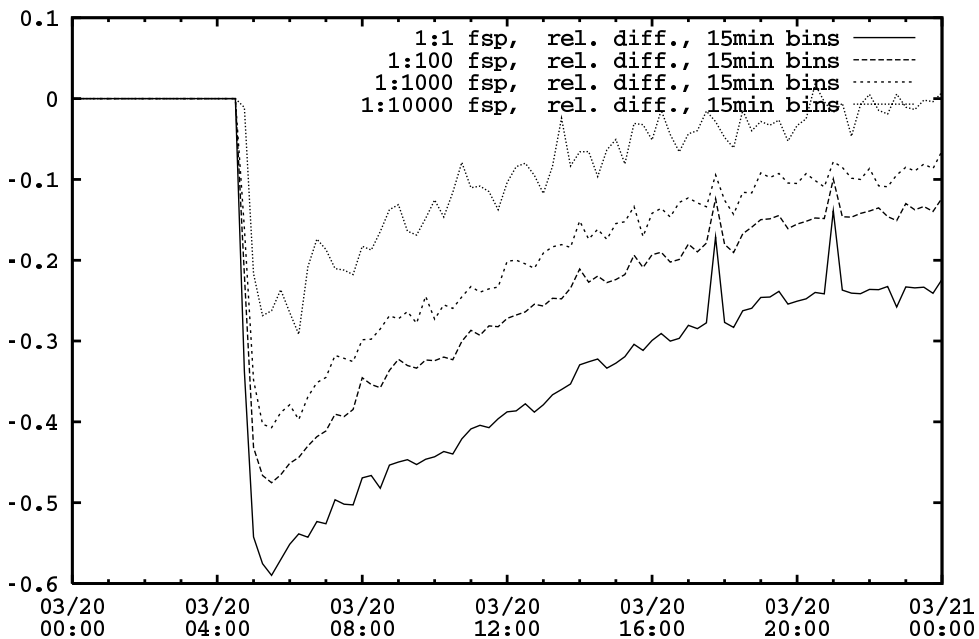




reldiff-20040320000-20040321000-9b - Relative Difference



reldiff-20040320000-20040321000-9b - Relative Difference



## C.4 Router 4

