A Measurement Study of the Origins of End-to-End Delay Variations

Yaron Schwartz, Yuval Shavitt, and Udi Weinsberg

School of Electrical Engineering
Tel-Aviv University, Israel

Abstract. The end-to-end (e2e) stability of Internet routing has been studied for over a decade, focusing on routes and delays. This paper presents a novel technique for uncovering the origins of delay variations by measuring the overlap between delay distribution of probed routes, and how these are affected by route stability.
Evaluation is performed using two large scale experiments from 2006 and 2009, each measuring between more than 100 broadly distributed vantage points. Our main finding is that in both years, about 70% of the measured source-destination pairs and roughly 95% of the academic pairs, have delay variations mostly within the routes, while only 15-20% of the pairs and less than 5% of the academic pairs witness a clear difference between the delays of different routes.

1 Introduction

The Internet has evolved in recent years to become a complex network, with increasing usage of load-balancing and traffic shaping devices. These devices change the way packets flow, therefore affecting the observed stability of routes and delays between hosts. This, in turn, affects various delay and jitter sensitive applications, such as VoIP and IPTV. On the other hand, load on devices is not constant and may change the delay packets observe along the same route significantly. Therefore, it is important to understand both the delay stability along the path and to identify the source of the delay variability when such variability exists.

Wang *et al.* [1] and more recently Pucha *et al.* [2] studied the impact that specific routing events have on the overall delay. They showed that although routing changes can result in significant round trip delay increase, their variability is small for most of the measured path transitions, therefore allowing applications to make use of such stability.

Augustin *et al.* [3] examined the delay between different parallel routes at a short time epoch. They compared the minimum delay of each route, and found that only 12% have a delay difference which is larger than 1ms. Using similar techniques, Pathak *et al.* [4] studied the delay asymmetry and found that there is a strong correlation between changes in the one-way delay and corresponding route changes.

Unlike previous work, we study the RTT delay along longer time periods, hours and days, and examine how different is the delay *distribution* between parallel routes. For this purpose we use delay samples to define an interval in which the delay of each route resides, and look at the overlapping between intervals of parallel routes. If the two intervals are disjoint we know that the e2e delay value mostly depends on the route in use and not on the variance in the route. As the overlap between the intervals increases, the delay variance is mostly attributed to changes along the route itself, e.g., due to change in load.

Evaluation is performed by conducting two large-scale experiments in 2006 and in 2009. Using DIMES [5], a highly distributed community-based measurements infrastructure, we planned these two 96-hours experiments each utilized more than 100 actively measuring vantage points (VPs), located in a broad set of ASes and geographical locations, contributing more than 200k e2e routes.

Our main finding is that in about 70% of the measured source-destination pairs, in both experiments, the delay variations are mainly explained by changes within the routes, while only 15-20% of the pairs witness a clear difference between the delays of different routes. The remaining 10-15% of the pairs witness a mixture of the above, with a higher tendency for intra-route changes as contributors to the delay variance. Pairs that have their source and destination in academic ASes exhibit much higher route stability, which further increases the percentage of delay variations within the routes to 95%.

2 Quantifying Route and Delay Stability

2.1 Definitions

The input data is a collection of traceroute measurements for a set P of ordered source destination pairs, $P_i = \{S_i, D_i\}$. For each pair, P_i, the set of e2e IP-level traceroutes, TR_i, is partitioned into k_i equivalence subsets (i.e., any two traceroutes in each subset are the same), denoted by E^i. The size of the subset $|E^i_j|$ is the total number of traceroutes it contains. Each equivalence subset $E^i_j, 1 \leq j \leq k_i$ has a single representing route $R(E^i_j)$ which is the measured path between the source and the destination.

For each pair P_i we define the *dominant route* as the route $R(E^i_j)$ whose subset size, $|E^i_j|$, is the largest. It is possible that several equivalence subsets have the same size, therefore they are all considered dominant routes. For brevity, we assume for now that each pair has a single dominant route, with index r.

2.2 Measurement Setup

The data used in this paper is obtained from DIMES [5], a community-based Internet measurements system. DIMES performs active measurements using hundreds of software agents installed on users' PCs. Agents perform roughly two measurements per minute (either traceroutes or ping using either ICMP or UDP) by following a script that is sent to them from a central server.

DIMES provides researchers with the ability to run "experiments" by defining the set of agents, probing protocols and a set of destinations. Since some agents are installed on end-users machines, the number of measurements may vary depending on their availability. Usually, more than 80% of the planned measurements are performed.

For the purpose of this paper we performed two similar experiments that took place in December 2006 and September 2009. In each experiment, we selected over 100 globally distributed agents and designed 96-hours experiments in which each agent executed UDP and ICMP traceroute measurements to all other agents in a round robin fashion. For each traceroute measurement we take the minimum delay of at most four probes sent over a period of a few seconds (in case of a lost probe we do not send another one instead). Since DIMES is a community based platform not all of the agents are constantly active during the experiments period. Moreover, since there is a certain churn in users along time, not the same agents were selected in both experiments. Thus, 120 agents were selected, making sure that there will be valid results from more than 100 agents. The scripts we wrote had one UDP and one ICMP measurement to each of the 120 destination IP addresses. Therefore, an agent probes each IP address twice every two hours. Agents repeated the same script for four days. In total, each of these experiments result in over one million traceroute measurements results.

Note that traceroutes probe the forward-path of routes, while the delays are round-trip. Pathak *et al.* [4] analyzed the delay asymmetry and showed that one-way delay can be different than round-trip, meaning that it is possible that our delay measurements actually capture instability that exists in both the forward and reverse paths. Following Pucha *et al.* [2], we analyzed the stability of routes as measured from opposite directions in our dataset, and found that over 90% of the pairs have forward and reverse path RouteISM that are different by less than 0.3 (not shown due to lack of space). This indicates that the stability of the forward path can serve as an indication to the reverse path. We attribute this to the observation that even non-symmetric routes share similar hops that can contribute instability to both directions. Thus, comparing the instability of RTT delay with the routing instability of the forward route is meaningful.

2.3 Pair and Route Identification

When comparing two routes we seek to answer if they are equal and if not, quantify their difference. Several difficulties arise in both aims. Since DIMES is a community-based project, most traceroutes start with several private IP addresses before reaching the routable Internet. Moreover, some use laptops and may travel during the time of the experiment. In order to decrease the chance of over-estimating instability, only the routable section of each traceroute is considered for the analysis. The identification of a pair is done using the first and last hops of the routable traceroute. This help us mitigate instability that might appear in the non-routable networks, which are presumed to have little affect on the overall delay instability. In the analysis, we only include pairs that witness at least 20 traceroutes.

Two (routable parts of) traceroutes are considered equal when their ordered list of IP addresses are exactly the same. To quantify the difference between two traceroutes we calculate their Edit Distance [6] (ED) value by counting the minimal number of insert, delete, and modify operations that are needed in order to make the two routes equal. Obviously, ED is highly correlated with the length of the compared routes. To be able to compare ED values that are calculated on routes with various lengths, the ED is normalized by the length of the longest route of the two input routes. This technique is similar to the one described by He *et al.* [7] who used it for quantifying AS-level asymmetry. We extend here the technique to consider stability instead of symmetry. Since the ED cannot be greater than the longest route, the normalized ED value is between 0 and 1, where 0 means that the two routes are identical and 1 means that they are completely different.

2.4 Route Stability

We use two methods for quantifying the stability of a route. The overall appearance ratio (i.e., prevalence [8]) of a route with index j, i.e., $R(E_j^i)$, in pair P_i is the portion of traceroutes in the set E_j^i. The prevalence of the dominant route $R(E_r^i)$ is used as the first indication to the stability of routing for each pair, since having a dominant route with high prevalence suggests that the remaining paths are relatively rare.

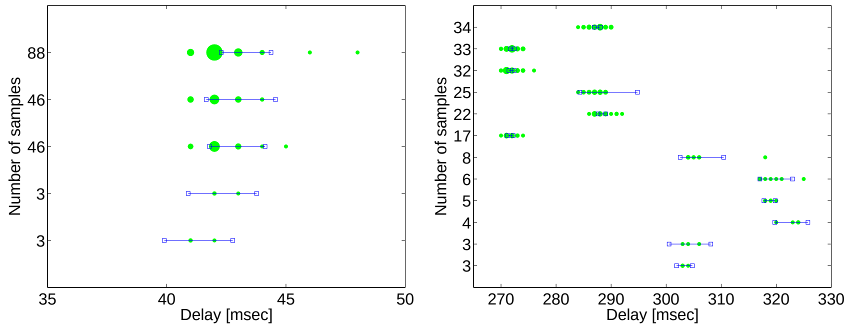
The second estimation of pair P_i stability is calculated by finding the normalized ED between the dominant route, $R(E_r^i)$, to all other non-dominant routes, $R(E_j^i), j \neq r$. For pairs that have more than a single dominant route, we use the dominant route that is closest to each route in number of hops. We define the *Route Instability Measure* (RouteISM) of a pair as the weighted average of all normalized ED measures as depicted in Eq. (1). Thus, an ISM value close to 1 indicates high instability.

$$RouteISM_i = \sum_{j \neq r} \left(|E_j^i| \cdot \widehat{ED}_{jr}^{\,i} \right) / \sum_{j \neq r} |E_j^i| \tag{1}$$

Two techniques were used in the past to measure distance between routes. Pucha *et al.* [2, 4] defined the similarity coefficient for calculating AS level route symmetry as the number of similar elements divided by the total number of distinct elements in the two routes $\frac{|P_i \cap P_j|}{|P_i \cup P_j|}$. He *et al.* [7] used string matching which is similar to our ED. We follow the latter and argue that ED better captures stability since it takes into account the *order* of elements in each route.

2.5 Delay Stability

We are interested in the expected e2e round trip delay of a route over time and not in short term congestion. Recall that we take the minimum delay of at most four probes sent over a period of a few seconds, and repeat each traceroute roughly twice an hour (UDP and ICMP) over a period of four days.



(a) 217.0.116.82 → 80.91.184.206 (b) 69.134.208.1 → 134.159.160.58

Fig. 1. Examples of pairs with overlapping and non-overlapping confidence intervals. The segments show the confidence intervals of a routes, calculated using the delay samples which are shown as varying sized circles (larger radius means more samples)

For a given pair P_i, each equivalence set E_j^i, has several different e2e RTT delay samples (henceforward "delays"), denoted by $RTT(E_j^i)$. We wish to quantify the stability of pair delays and whether their variance is the result of delay dynamics of each route or delay difference between different routes. This analysis can uncover whether delay instability is mainly the result of traffic anomalies in a route (e.g., congested routers), or the result of route diversity due to load-balancers.

For each route E_j^i, we have the group $RTT(E_j^i)$ of delay measurements. To find the region of expected delay for the route, we treat the measured delays as samples of some distribution and calculate the average and the 95% confidence interval [9] around it. This confidence interval, denoted by $CI(E_j^i)$, provides us with a segment surrounding the measured mean of $RTT(E_j^i)$. Within this interval we expect to find the route delay. Note that this is an unorthodox use of confidence interval, but we believe it gives us a good characterization of the expected route delay (as is nicely shown in Fig. 1). Measurements with high variance result in larger segments than measurements with small variance, indicating that they are less stable.

For a source-destination pair, the normalized overlap between two segments $CI(E_j^i)$ and $CI(E_k^i)$ is defined by

$$\widehat{O}_{jk}^i = \frac{CI(E_j^i) \bigcap CI(E_j^i)}{\min\{|CI(E_j^i)|, |CI(E_j^i)|\}}, \forall j \neq k \tag{2}$$

The normalized overlap is equal to 0 when the two segments do not overlap, meaning that their delays are significantly different. This indicates that changes in the route delay are mainly the result of having different routes. When it is equal to 1, the segments completely overlap or one contains the other, meaning that different routes exhibit similar delay distribution, indicating that instability is not the result of multiple routes between the source and destination, but due to changes within the routes. For example, Fig. 1(a) shows the routes from Deutsche Telekom in Germany to Datagroup in the Ukraine. There are five different routes with more than 30 measurements (the y-axis label is the number of measurements per route) but they are all overlapping, namely they have roughly the same delay average. On the other hand, Fig. 1(b) shows the routes from ParaCom Technologies in USA to Reach Networks in Australia. While in the previous figure the delay changes are attributed to variance of delays inside the routes, this figure clearly shows that the delay changes is the result of multiple routes with four distinct mean delays.

When the number of measurements in the route is small, the statistical significance of the samples is small, and the confidence interval can be very large and not meaningful. Fig. 4(a) shows that for 80% of the routes the confidence interval is below 0.2 of the average delay. Since the number of routes with statistical significance change between pairs, we calculated the overlap only between the two largest equivalence groups (routes) of each pair, providing each has at least 30 delay measurements.

3 Dataset Analysis

3.1 Distribution of Vantage Points

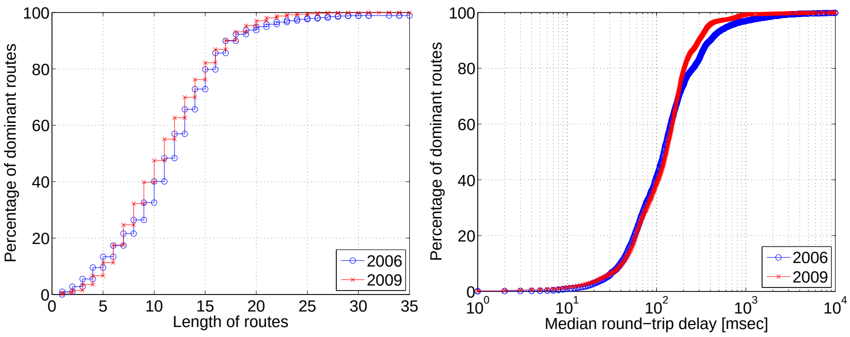
Using a community based platform, results in a certain churn in the availability of measuring agents. Therefore, during the planning of the experiments, we selected measuring agents that hold all of the following criteria: (a) they were active in the past week, (b) distributed in a large set of ASes, and (c) distributed in a large set of geographical regions. The first criterion is to maximize the chance that the selected agent will indeed be active during the experiment period. The other two criteria were selected to achieve e2e routes with diverse lengths that traverse through various ASes spread across different countries and continents, as an attempt to capture an accurate image of the Internet [10].

In the 2006 experiment, 102 agents returned slightly over a million traceroutes, providing us with 6861 source-destination directed pairs. Most VPs are distributed in the USA and Canada (70), followed by Western Europe (14), Australia and New Zealand (10), Russia (6) and Israel (2). In the 2009 experiment, 105 agents returned 1.01 million traceroutes, resulting in almost 10950 source-destination directed pairs. VPs are distributed in numerous countries in Western Europe (41), followed by USA and Canada (38), Russia and the Ukraine (14), Australia (4), South America (2), Israel (2), Japan (1), Taiwan(1), Singapore (1) and the Maldives (1).

Using the list of AS types provided by Dimitropoulos *et al.* [11], we infer the type of each VP. In 2006, 18% of the VPs are tier-1, 78% tier-2, 3% smaller companies and 1% educational (a single agent). In 2009, DIMES agents were installed in PlanetLab [12], which increased the number of educational VPs to 28% while reducing tier-1 VPs to 14% and tier-2 to 58%. Only 7 VPs appeared in both experiments. This is due to the change in users that are running DIMES agents over this time period.

In both experiments, a variety of ASes were traversed. Most of the tier-1 ASes were traversed and the majority of the traversed ASes are tier-2.

3.2 Dataset Statistics

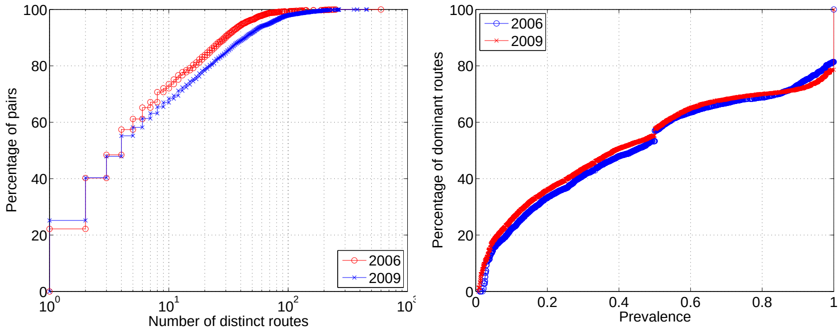


(a) Route Length (b) Median Delay

Fig. 2. Cumulative distributions of route lengths and median delay

The cumulative distributions of the dominant route length and dominant route median delay are shown in Fig. 2 (recall that there can be more than one dominant route per source-destination pair). Fig. 2(a) shows that both experiments have roughly the same path lengths, with 2009 being slightly shorter. The median of the dominant route length is 12 for 2006 and 11 for 2009; pairs with academic source and destination ASes have even shorter routes, with median of 11 hops in 2009; the majority of the routes (97%) traverse less than 20 hops. Our measured routes are shorter than reported by Paxson [8] in 1995. Paxson reported mean route length between 15 and 16, using routable (and mostly academic) source and destination hosts. Since the Internet has been growing at high rate since 1995, we attribute this reduction to the reacher connectivity among ASes and increased adoption of layer-2 tunnels, which significantly reduces the number of IP-level hops.

Fig. 2(b) exhibits an almost identical median delay distribution of 2006 and 2009, with 2009 having slightly shorter delays, which correlates with the shorter paths witnessed in Fig. 2(a). Over 80% of the routes in both years have a delay of less than 200msec. However, there are almost 3% of the routes that have a delay of over 1 second. Pairs that have end-points in the USA are have shorter delays, with 80% of them having a delay less than 150msec. However, pairs with academic end-points have significantly shorter delays, with 90% of them having a delay of less than 100msec.



(a) Distinct routes per pair (b) Dominant route prevalence

Fig. 3. Cumulative distributions of distinct routes and prevalence, showing (a) the number of distinct routes per pair, and (b) the prevalence of the dominant route

4 Results

4.1 Route Stability

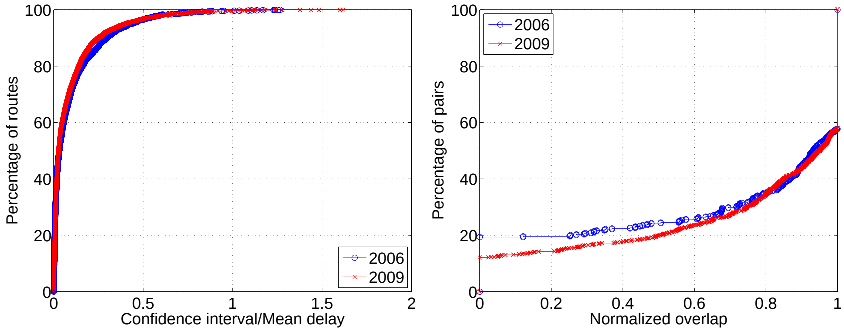
Fig. 3 shows the cumulative distributions of distinct routes per pair and prevalence of dominant routes. The figures show that over 20% of the pairs in 2006 and almost 30% of the pairs in 2009 witness a single route. Fig. 3(b) has a clear jump at 50% prevalence, which we attribute to load-balanced routes with equal per-packet balancing. This jump is not visible in routes between academic end-points, due to their minimal usage of load-balancers. Furthermore, over 55% of the pairs that have both source and destination in academic ASes, which is the case when using PlanetLab, have a single route. Pairs that have both end-points in the USA have slightly higher route stability, with roughly 35% of them having a single route. These observations stress the need for a diverse set of VPs when doing e2e Internet analysis.

Analysis of the RouteISM (not shown due to lack of space) supports the observation of an overall stable e2e routing in the Internet, as over 90% of the pairs (and 95% of the academic pairs) have RouteISM smaller than 0.2. This values is used in Sec. 4.2 as a threshold between stable and non-stable pairs.

4.2 Origin of Delay Instability

We first show that our use of confidence interval is meaningful. Fig. 4(a) plots the cumulative distribution of the ratio between a route's confidence interval and its mean delay. The figure shows that, for both years, 90% of the routes have a ratio of less than 0.25. This indicates that the delay confidence intervals are not 'too long' in general, and extend only for routes with large variance (as shown in the examples in Fig. 1).

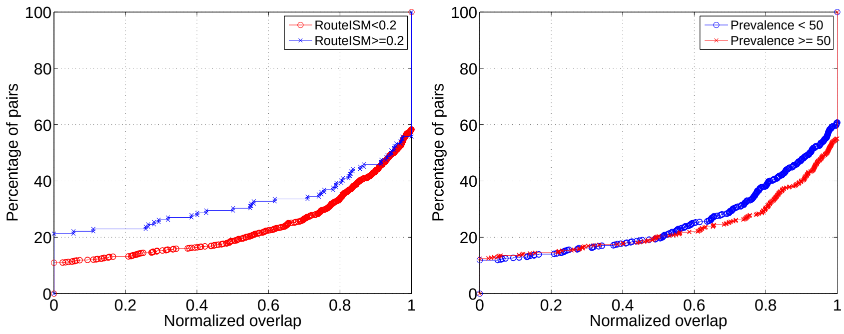
Fig. 4(b) shows that for both data sets, over 40% of the pairs have an overlap of 1 and an additional 30% of the routes have overlap of over 0.8. Namely, in 70%



(a) Confidence interval/mean delay (b) Normalized overlap

Fig. 4. Confidence interval statistics

of the cases changes in route delay cannot be attributed to multiple path routing but rather to changes between the routes. In 15% of the cases (20% in the 2006 data sets) the change in delay is mainly due to route changes as the overlap is zero or close to 0. Over 95% of the pairs that have academic source and destination ASes have an overlap of over 0.7. This is mainly the result of academic networks having small routes difference (induced by local load-balancing) and little usage of "spill-over" backup routes. Only 5% of the pairs that have both source and destination in the USA witnessed overlap of 0.



(a) RouteISM (b) Prevalence

Fig. 5. Effect of route stability on normalized overlap

Finally, we evaluate how the route stability affects the overlap of delays. Fig. 5(a) shows that routes with high RouteISM (≥ 0.2) have higher percentage of non-overlap delay intervals. Namely, when the difference between the routes is larger, there are higher chances that their delay distribution will be different.

Fig. 5(b) shows that, unlike RouteISM, the prevalence of the dominant route does not significantly affect the level of overlap.

5 Conclusion

This work presents a measurement study of the e2e delay variance and its origins. Given a set of probed RTT delays, we find a confidence interval which better captures the delay of each observed route. We then compute the overlap of these intervals for uncovering the origin of these variations. Additionally, we develop techniques for quantifying route stability and measure its affect on the origin of delay variance. We find that for roughly 70% of the pairs and for over 95% of the academic pairs, the delay variations are mostly within the routes and not between different routes.

Acknowledgment. This work was partially funded by the OneLab II and the MOMENT consortia that are part of the EU FP7; and the Israeli Science Foundation, grant 1685/07.

References

1. Wang, F., Mao, Z.M., Wang, J., Gao, L., Bush, R.: A measurement study on the impact of routing events on end-to-end Internet path performance. ACM SIGCOMM CCR **36**(4) (2006) 375–386
2. Pucha, H., Zhang, Y., Mao, Z.M., Hu, Y.C.: Understanding network delay changes caused by routing events. In: SIGMETRICS. Volume 35. (2007) 73–84
3. Augustin, B., Friedman, T., Teixeira, R.: Measuring load-balanced paths in the Internet. In: IMC. (2007)
4. Pathak, A., Pucha, H., Zhang, Y., Mao, Z.M., Hu, Y.C.: A Measurement Study of Internet Delay Asymmetry. In: PAM. (2008)
5. Shavitt, Y., Shir, E.: DIMES: Let the internet measure itself. ACM SIGCOMM CCR **35**(5) (2005) 71–74
6. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady **10**(8) (1966) 707–710
7. He, Y., Faloutsos, M., Krishnamurthy, S.: Quantifying routing asymmetry in the internet at the AS level. In: GLOBECOMM. (2004)
8. Paxson, V.: End-to-End Routing Behavior in the Internet. In: IEEE/ACM Transactions on Networking. (1996) 601–615
9. Bolle, R.M., Ratha, N.K., Pankanti, S.: An evaluation of error confidence interval estimation methods. International Conference on Pattern Recognition **3** (2004)
10. Shavitt, Y., Weinsberg, U.: Quantifying the importance of vantage point distribution in Internet topology measurements. In: Infocom. (2009)
11. Dimitropoulos, X., Krioukov, D., Riley, G., kc claffy: Revealing the AS taxonomy: The machine learning approach. In: PAM. (2006)
12. Chun, B., Culler, D., Roscoe, T., Bavier, A., Peterson, L., Wawrzoniak, M., Bowman, M.: Planetlab: An overlay testbed for broad-coverage services. ACM SIGCOMM CCR **33**(3) (July 2003)