UNIVERSITY OF CAPE TOWN
DEPARTMENT OF COMPUTER

COMPUTER SCIENCE HONOURS
FINAL PAPER
2015

Title: Topology Discovery for African NRENs

Author: Roslyn Sanby

Project Abbreviation: AfriNREN

Supervisor: Assoc. Prof. Hussein Suleman and Josiah Chavula

Category	**Min**	**Max**	**Chosen**
Requirement Analysis and Design	0	20	5
Theoretical Analysis	0	25	0
Experiment Design and Execution	0	20	20
System Development and Implementation	0	15	0
Results, Findings and Conclusion	10	20	20
Aim Formulation and Background Work	10	15	15
Quality of Paper Writing and Presentation	10		10
Adherence to Project Proposal and Quality of Deliverables	10		10
Overall General Project Evaluation (*this section allowed only with motivation letter from supervisor*)	0	10	
Total marks		**80**	**80**

Topology Discovery for African NRENs

Roslyn Sanby
University of Cape Town
Rondebosch, Cape Town
7700
+2721 650 9111
sanbyra@gmail.com

ABSTRACT

National Research and Education Networks (NRENs) aim to provide IP connectivity between educational institutions to improve education and research. In Africa, about 75% of the network traffic between African NRENs follow circuitous routes through Europe and other areas outside the continent. In order to find out what the extent of these circuitous routes are, research needs to be done into discovering the topology of the Internet in Africa and how traffic is being exchanged between and within the NRENs in Africa. Conducting Traceroute measurements from multiple vantage points in a variety of locations helps with limiting the bias in the topology and gives more complete data. Data can be acquired more efficiently by finding overlapping paths and thus reducing redundancies in measurements. This data can then be visualised for future research into the effect of these circuitous paths. The visualisations can also help network managers plan routing policies to reduce latencies.

CCS Concepts

Networks → Network components,

Network → Network services → Network monitoring,

Network → Network services → Network management,

General and reference → Cross-computing tools and techniques → Measurement,

General and reference → Cross-computing tools and techniques → Evaluation

Keywords

Networking; NRENs; Traceroute; RIPE Atlas; latency; UbuntuNet Alliance; Africa; path diversity; topology

1. INTRODUCTION

A National Research and Education Network (NREN) is a mesh of interconnected networks that support the needs of education and research communities in a country [7]. The UbuntuNet Alliance is an association of NRENs in Southern and Eastern Africa that, as of September 2014, consisted of 15 NRENs [7].

This thesis was completed as part of a Bachelor of Science Honours degree in Computer Science at the University of Cape Town in 2015. For more information about this project, please see http://people.cs.uct.ac.za/~snbros001/AfriNRENWeb2/.

One of the aims of an NREN is to reduce latencies between educational institutions in order to facilitate better research and communication [7]. Other continental NRENs, such as GÉANT in Europe and RedCLARA in Latin America, have achieved these aims. However, the objective of reducing latencies between African NRENs faces challenges. It has been found, for example, that about 75% of the traffic between NRENs in Africa is routed outside the continent, travelling circuitous routes, often through Europe [3] [9]. This results in high latencies, which could be avoided by introducing more direct physical links or local peering [3] [9].

There is a need for a platform which collects and displays accurate data about NREN topologies in Africa. There is currently no similar work being done to map and analyse the topologies of African NRENS and to see where latencies could be reduced by avoiding circuitous intercontinental routes. A platform that visualises accurate topological data as well as latencies experienced by NRENs will help future researchers to see which networks exchange data circuitously and evaluate how and where networks can be improved. It will also add to the discussion and argument for why more local peering could be beneficial to African networks.

This paper aims to find out: can Traceroute data be collected reliably and efficiently for the purpose of discovering the topology of African NRENs? By "reliably", it is meant to increase the accuracy of these measurements and ensure that the topology discovered is complete. A topology is complete if all the paths defining that topology have been found. By "efficiently", it is meant to reduce the number of measurements to perform for obtaining a complete topology. A reduction in packets sent would accomplish this goal. This study intends to answer this question by conducting various Traceroute measurements from twelve vantage points to fifty destinations as outlined in Section 3.

2. BACKGROUND

2.1 National Research and Education Networks (NRENs)

The goals of an NREN, apart from providing IP connectivity, include providing point-to-point connectivity or bandwidth-on-demand [7]. This is done by taking advantage of Internet Exchange Points (IXPs) and peerings to improve the communication between educational institutions. Internet Exchange Points (IXPs) are infrastructures where Internet traffic is exchanged, allowing networks to connect directly and reduce latency, bandwidth and cost [9]. However, it has been found that about 75% of traffic originating in Africa and destined for other African NRENs travels out of the continent to reach their destination [3].

Chavula et al. [3] and Gupta et al. [9] have done research on Internet traffic in Africa travelling intercontinentally and found that these circuitous paths result in higher latencies. Gupta et al. [9] focused their research on the connectivity at JINX in Johannesburg and

KIXP in Nairobi, which are major IXPs in Africa, while Chavula et al. [3] researched NRENs in Africa, gathering information from five vantage points targeting 95 universities. These slightly different research strategies seem to reach the same conclusion that NRENs in Africa, as a result of circuitous Internet paths, are not meeting their objective of reduced latencies.

Gupta et al. [9] give two suggestions to avoid these circuitous paths and thus reduce latencies. One is to add more local cache servers in Africa. If one looks at content hosted overseas, such as Google or BBC, more local cache servers could help reduce latencies. However this approach is not necessarily beneficial for local content already hosted in Africa. An example of this content would be inter-university virtual classrooms where students learn online [10]. Direct local links to the cache servers would also be required for the data not to be exchanged circuitously [9]. The second suggestion is to add more peering relationships in Africa. This suggestion could be helpful, as it would create more local links and help NRENs avoid circuitous routes [18]. However, research needs to be done on where these peering relationships are needed and IXPs need to be encouraged to participate in the creation of more links. More research needs to be done on the level of traffic that is being exchanged between these networks [3] so as to determine best interconnection design.

2.2 Internet Topology Discovery

The discovery of Internet topologies and the monitoring of Internet traffic is a highly researched field. Studies have been conducted to attempt to map the Internet so as to gain a better understanding of where Internet traffic is routed and how latencies, bandwidth and other metrics could be improved [8] [4].

Measurements of the Internet can be performed in the control or data planes. Measurements in the control plane consist of information about Internet routing, often found in Border Gateway Protocol (BGP) tables [14]. Measurements in the data plane look at which paths packets travel along to discover the reachability of the Internet as well as Round Trip Time (RTT) of packets [14]. These measurements can be either active or passive. The difference between active and passive measurements is that active measurements tools, such as Traceroute, send packets into the network and collect data from the response. Passive measurements, like BGP monitors, collect information that is already flowing over the wire [14]. Active measurements are often used for collecting data to discover topologies whereas passive measurements are used for traffic profiling.

Topology information can be collected on four different levels of granularity – Internet Protocol (IP) Interface, Router, Point of Presence (PoP) and Autonomous System (AS) levels [6] [14]. NRENs mainly operate on the AS level but could constitute several PoPs. Topology discovery at the PoP level provides information and limitations about latencies between PoPs. This helps with understanding the geographical properties of Internet paths, such as where ASes can connect and the coverage of ASes [14]. This project focused on methods of discovery at the AS and PoP levels.

Autonomous Systems (ASes) are privately managed networks, which are all interconnected making up the Internet [6] [14]. ASes are identified by a unique 16-bit AS number. To collect information for the AS level, data is collected from BGP tables, Traceroute measurements and Internet Routing Registries (IRR). This indicates both active and passive measurements in both the control and data plane.

A Point of Presence (PoP) is a collection of routers belonging to one AS [14]. There are three main methods to collect data at the PoP level. Firstly, data can be aggregated from Traceroute measurements to identify PoPs. Secondly, delay estimates can be obtained from Ping measurements. Finally, information can be retrieved from websites where ISPs have published their data. Although this method will provide more accurate data than the active measurements, the technique is not always reliable as the information could be outdated [14].

Traceroute appears to be the best method for probing the Internet to gather information about its topology and how traffic is routed between and within networks [4] [14] [13] [16] [2]. There are three main variants of Traceroute which make use of different protocols. The standard Traceroute uses User Datagram Protocol (UDP) probes and receives Internet Control Message Protocol (ICMP) responses. Another variant uses only ICMP by sending ICMP echo requests and receiving ICMP echo replies. These two variants, however, can encounter errors if a router does not have the ICMP protocol enabled or if a router employs ICMP rate limiting [6]. The third variant makes use of Transport Control Protocol (TCP) packets and sends TCP SYN packets to try to get past the most common firewall filters [6].

One further variant of Traceroute is Paris Traceroute. Paris Traceroute is made up of ICMP-Paris and UDP-Paris [13] and helps in discovering alternate paths. It avoids missing links and nodes as well as false links which could appear because of load balancing [2] [6]. This is done by controlling and varying the packet header contents when conducting Traceroute measurements.

2.3 Distributed Network Probing

Distributed network probing uses many vantage points to get a more accurate view of the Internet [18] [19]. The use of many vantage points when conducting Traceroute measurements helps to provide a more accurate view of a topology. Varying the location of those vantage points further increases the accuracy [19]. More vantage points reveal new links between ASes [18]. However, if the vantage point is situated outside the two local Internet Service Providers (ISPs) in question, a connection may be missed due to BGP policies [18].

There are many platforms available to aid with distributed probing. Some examples are Archipelago [11] [5] [12], DIMES [18], iPlane [19] and RIPE Atlas [17].

RIPE (*Réseaux* IP *Européens*) Atlas is a platform that makes use of thousands of active probes around the world to measure Internet connectivity and reachability in real-time [17]. These probes are small USB-powered hardware devices (attached to an Ethernet port) that conduct measurements, such as Ping, Traceroute, DNS and SSLcert, and relays the data to the RIPE Network Coordination Centre (NCC). This data is then aggregated with data collected from other RIPE Atlas probes. The RIPE Atlas platform is very advantageous for researching the topology of African NRENs. Out of the platforms available, RIPE Atlas has the most probes situated in Africa [17]. Custom measurements can be sent from a probe to any IP address, and therefore to any NREN, allowing the collection of data that is required to study routes between African NRENs.

Of the platforms available, RIPE Atlas was chosen for this project because it was found that it has the most vantage points in Africa and specifically in NRENs in Africa.

2.4 Other Related Work

Chavula et al. [3] used the CAIDA (Centre for Applied Internet Data Analysis) Archipelago platform to conduct their experiments. This made use of Scamper to collect data about intra-Africa Internet traffic. Scamper is a tool that utilises four different probing techniques. One of these probing techniques was Paris Traceroute. Gupta et al. [9] used Traceroute as well as BGP routing tables for collecting their data.

Aben used the RIPE Atlas platform to measure and analyse Internet traffic paths particularly in Sweden [1]. His research method was similar to that of Chavula, et al. [3] as he used multiple vantage points to get a more accurate view of the topology of the Internet. Aben [1] had access to hundreds of probes around Europe to investigate traffic traversing IXPs. However, there are only 19 RIPE Atlas probes within NRENs in the UbuntuNet Alliance. Chavula, et al. [3] used five vantage points because the CAIDA Archipelago platform [11] [5] [12] only has five vantage points in Africa.

3. EXPERIMENTAL DESIGN

The design goals are to obtain reliable and accurate results efficiently. This requires running multiple Traceroute measurements using different protocols, from different vantage points, to a diverse set of targets located in NRENs in the UbuntuNet Alliance. The paths need to be analysed in order to find overlaps and reduce redundancy in further measurements.

The data collected is needed for NREN CEOs and managers to discover patterns and determine where and why latencies are occurring. In this way NREN CEOs and managers could see where the performance of their NRENs could be improved. This data should be accurate and complete so that a useful visualisation can be produced.

3.1 Requirements Gathering

Surveys and interviews were conducted to gather requirements. These indicated which data, when collected and analyzed, would result in a useful and comprehensive visualization. Questions were asked about limitations in network management and current network management software, among others.

After receiving ethical clearance, a survey was sent out to network managers and CEOs of NRENs in the UbuntuNet Alliance. One comment from the survey was that a common network management limitation is a lack of comprehensive routing information.

An interview with a member of UCT's Information and Communication Technology Services (ICTS) highlighted the tools already available for network management. These tools showed to some extent what data would be required for an effective visualisation, one which shows how data is transmitted between NRENs. Probes and IP address targets need to be chosen within NRENs in the UbuntuNet Alliance. The Round Trip Times are required for each hop to indicate the latencies between hops. Coordinates and AS numbers are also needed for the hop and target IP addresses so that they can be mapped for the visualisation.

3.2 Constraints and Considerations

The RIPE Atlas platform only has probes in a limited number of NRENs within the UbuntuNet Alliance. As a result, it is not possible to discover a complete topology of African NRENs using the RIPE Atlas platform, since there is not full coverage of vantage points in all NRENs. The aim, therefore, is to get the most complete topology using the probes available. Twelve probes were used, where at most two were selected from each AS identifying an NREN or university within an NREN.

Credits are required to run Traceroute measurements from the RIPE Atlas platform. Credits are a form of currency on the RIPE Atlas platform and are earned through hosting a probe and allowing it to be used for measurements by the public [17]. A probe is currently hosted at the University of Cape Town and therefore credits are available to conduct measurements.

3.3 Approach

3.3.1 Experiments to conduct

The RIPE Atlas platform allows for measurements to be conducted from any vantage point given that the probe at that point is active. For this design, at most two probes were selected from each NREN so as to spread out the selection of probes within the UbuntuNet Alliance. The ASes in which RIPE Atlas probes were found are shown in Table 1.

Table 1. List of AS numbers and their names

Name	AS number	Country
RENU	327687	Uganda
KENET	36914	Kenya
iRENALA	37054	Madagascar
SudRen	37197, 33788	Sudan
TENET	2018	South Africa
University of Cape Town	36982	South Africa
Rhodes University	37520	South Africa

The platform also allows for measurements to be sent to any target IP address. Fifty IP addresses were chosen. Each of these IP addresses represents a university or educational organisation within an NREN in the UbuntuNet Alliance.

The probes and the destinations used for the Traceroute measurements are depicted in Figure 1. Probes are represented by blue diamonds while destinations are represented by circles. The colour of the circle indicates in which AS the IP address is located.

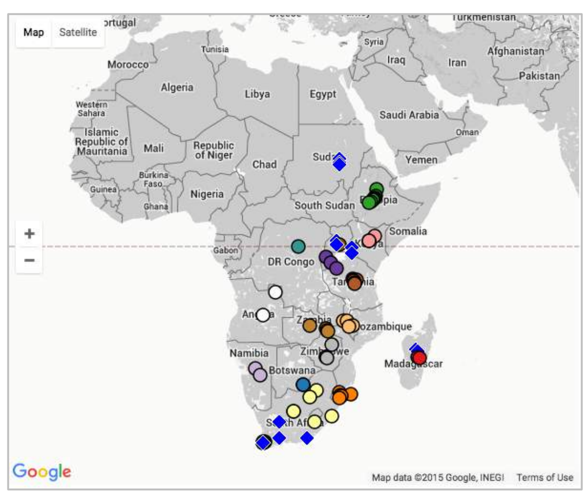

Figure 1. Visualisation of probes and destinations for Traceroute measurements

A set of source-destination pairs was constructed so that measurements could be conducted from each probe to each destination.

Traceroute measurements can be run using different protocols. The protocols that were used were ICMP, TCP and UDP. The goal was to discover different paths; since some protocols are blocked by firewalls on routers, several protocols were used.

A field can be edited when setting up the Traceroute measurement that changes the number of variations to be used for a Paris Traceroute. Paris Traceroute is less likely to experience anomalies as a result of load balancing. This field is set as a number between 1 and 64. Paris values of 0, 16 and 64 were used in this project. A Paris value of 0 means that a standard Traceroute is performed. Comparing the results from these measurements, it was found that there was no statistically significant difference between paths or RTTs when changing the Paris value. Therefore, in further measurements, a default value of 0 was used.

In order to ensure efficiency in performing measurements, overlapping paths had to be identified and accounted for. If the beginning of paths of two Traceroute measurements overlapped, one of the Traceroute measurements was configured to skip the first hops that were overlapping. If the ends of two paths overlapped, one of the Traceroute measurements was set up to trace only up to the hop of intersection along the path of the original Traceroute.

The experiment was initially run three times in an attempt to find as may paths as possible from each source to each destination. Routing and load balancing can provide a diverse set of paths depending on the congestion of the network and how packets were subsequently routed at a particular time, which is why the experiment was run more than once. Three full measurements were conducted from each probe to each target IP address using each of the three protocols. The paths of these measurements were analysed to find overlaps. Once this had been completed, three more sets of Traceroute measurements were run for each protocol where first hops and last hops were changed where necessary.

3.3.2 Tools designed

A set of Python libraries was used to perform most of the functions needed to conduct the Traceroute measurements [20]. These libraries use the RIPE APIs (Application Programming Interfaces) to perform functions including fetching a list of active probes and conducting Traceroute measurements.

After using a library to fetch all the probes that were active on the RIPE Atlas website, Awk scripts were used for text manipulation to select the specific probes within the ASes (shown in Table 1). At most two probes from each AS were chosen so as to get an even distribution of vantage points within the Alliance.

The Traceroute measurements were then conducted by using another of the Python libraries in the set mentioned above. The API for specifying User Defined Measurements on the RIPE Atlas platform is a RESTful API (Representational State Transfer Application Programming Interface). This API is used to set up a Traceroute measurement with specified settings from a probe to a destination. The settings that were used are presented in Table 2. Each measurement is referenced by its measurement identification (ID) number. The ID numbers for each measurement were stored so that results could be collected after the Traceroute measurements were completed.

Table 2. Settings for Traceroute measurements

Variable	Setting
Protocol	ICMP/UDP/TCP
Paris	0
Number of packets	3
Number of repeats of measurement	1

Each measurement was run three times, once with each of the three protocols. In an attempt to minimise packet loss, multiple packets can be sent in a Traceroute measurement. The default of 3 packets was used. The measurement can also be specified as a one-off measurement which delivers the results very quickly. To specify the measurement as a one-off measurement, the number of repeats was left as 1.

Once the Traceroute measurement was completed, the results were collected by using cURL, a Linux command line tool for transferring data, to access the measurement APIs on the RIPE Atlas website. This returned a JSON file with all the Traceroute information.

The results of these full measurements were analysed to find overlapping paths. Paths which overlapped at the beginning were found by creating a tree with the probe IP address as the root. As shown in Figure 2, each subsequent node was the IP address of the next hop.

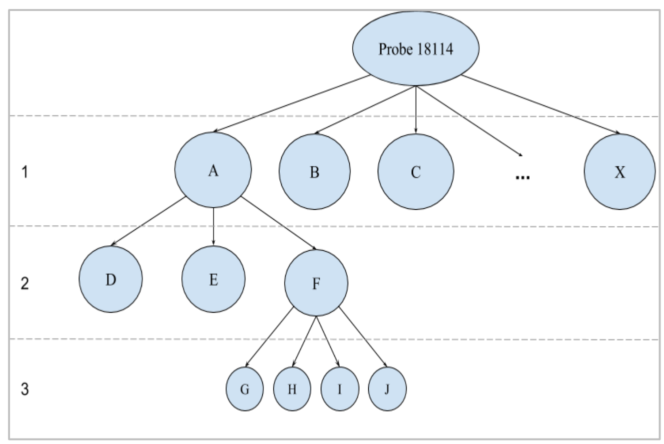

Figure 2. Example n-ary tree created to find overlapping paths at beginning of Traceroute measurements

If a probe followed the same path at the beginning of the Traceroute, the first hop field could be incremented when conducting the Traceroute measurements. This would exclude the overlapping path segment in further measurements. In the example in Figure 2, measurements from probe 18114 going to G, H, I and J all went through A and F. So, the first hop for G, for example, could be incremented to 2 (the last common hop). A record was made of the measurement which had the same beginning path to allow for reconstruction of the path information.

Paths which overlapped near the destination were found by comparing the paths to a particular destination. Paths from different probes to the same destination were compared. If the same IP addresses were found at the end of the path for two or more measurements from different probes, the measurement was set up to trace to the first hop of the intersection. For example, as depicted in Figure 3, if the measurements from X and Y went through both C and B to get to destination A, the Traceroute would be sent to C instead of A. A record was also made of the measurement which had the same end path.

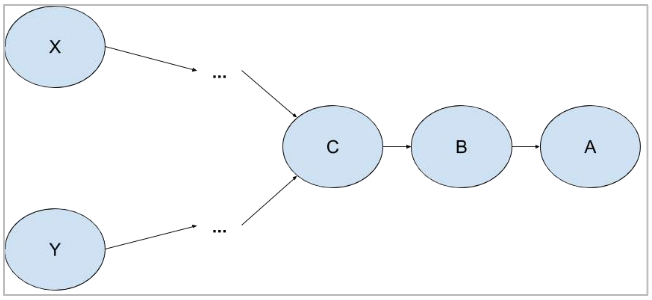

Figure 3. Example of overlapping paths at the end of two Traceroute measurements

In this way, subsequent measurements were conducted for each source-destination pair where first and/or last hops were altered where necessary. Some Traceroute measurements were left unmodified to allow comparisons with the whole path. Full path information could then be reconstructed by referencing the other measurements which had the same beginning or end paths.

Python scripts were used to aggregate the results. The average of the Round Trip Times from the three packets was set as the Round Trip Time for each hop. The latency of a measurement was set as the Round Trip Time of the last hop.

The output for each measurement was a JSON file, which was stored in a database for analysis and use in the visualisations. The geographical coordinates and AS number of each hop IP address and the target IP address were added to the JSON file by querying the Maxmind Geolite database of IP addresses.

Python scripts were also written for analysis and comparison of the results. Information such as percentage of source-destination pairs that reached a destination, diversity of paths found and the variance of latencies were calculated.

In summary, eighteen measurements were conducted for each source-destination pair. These measurements consisted of six ICMP-based measurements, six TCP-based measurements and six UDP-based measurements. The six measurements for each protocol consisted of the three full measurements and the three partial measurements where either the first hop or last hop was changed, or both, because of overlapping paths. In total, three sets of full measurements and three sets of partial measurements were conducted.

After these measurements were conducted, it was found that not all the probes had responded at the time the measurement was conducted. This was probably because the probe was offline at the time of the measurement. Therefore, more measurements were conducted to complete the set of eighteen measurements per source-destination pair.

4. RESULTS

4.1 Destinations Reached

As seen in Figure 4, 50% of the measurements configured with ICMP reached a destination. About 63% of the TCP-based measurements reached a destination. Only 26% of the UDP-based measurements reached a destination.

Therefore, out of the three protocols, measurements configured with TCP reached the destination most often. This clearly indicates that measurements configured with TCP had the highest reachability in terms of number of measurements which reached a destination.

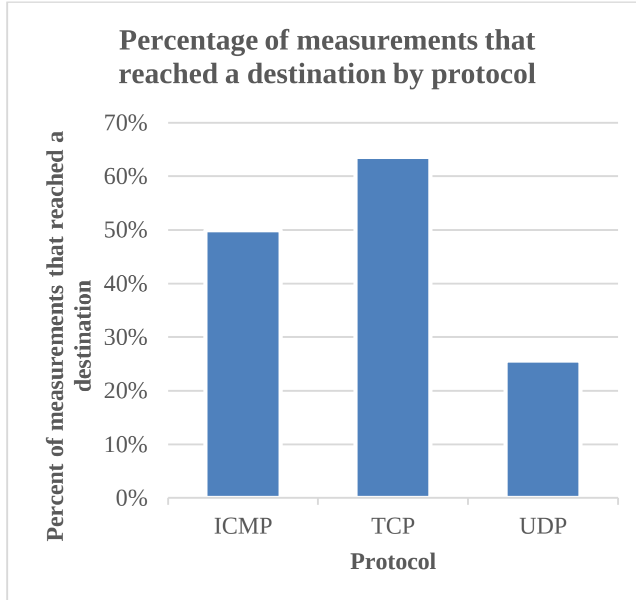

Figure 4. Graph showing spread of measurements that reached their destination by protocol as a percentage of all measurements

Out of the fifty target destinations that were probed, all of the destinations were reached by at least one probe. TCP-based measurements performed the best by reaching 100% of the destinations. ICMP-based measurements reached 70% of the destinations and UDP-based measurements reached 56% of the destinations. This was to be expected, since ICMP and UDP-based measurements are often blocked by firewalls as mentioned in Section 3.3.1. Therefore, in terms of reachability, there does not seem to be an advantage to using other protocols besides TCP. TCP-based measurements reached all of the destinations, therefore no other protocol is required to reach any missing destinations.

4.2 Path Diversity

When there is more than one unique path from a source to a destination, there is path diversity. For this section, only full measurements are considered.

As seen in Figure 5 below, more than half of the source-destination pairs have more than one unique path that reach the destination. In fact, 78.9% of these pairs have more than one path.

There are 45 (8%) source-destination pairs where the number of unique paths is zero – for example, probe 18149 to the first destination 137.158.158.44. This is because there are no paths from that source to that destination which reach the destination. There are incomplete paths from the source to the destination where the last hop does not respond, but none that reach the destination.

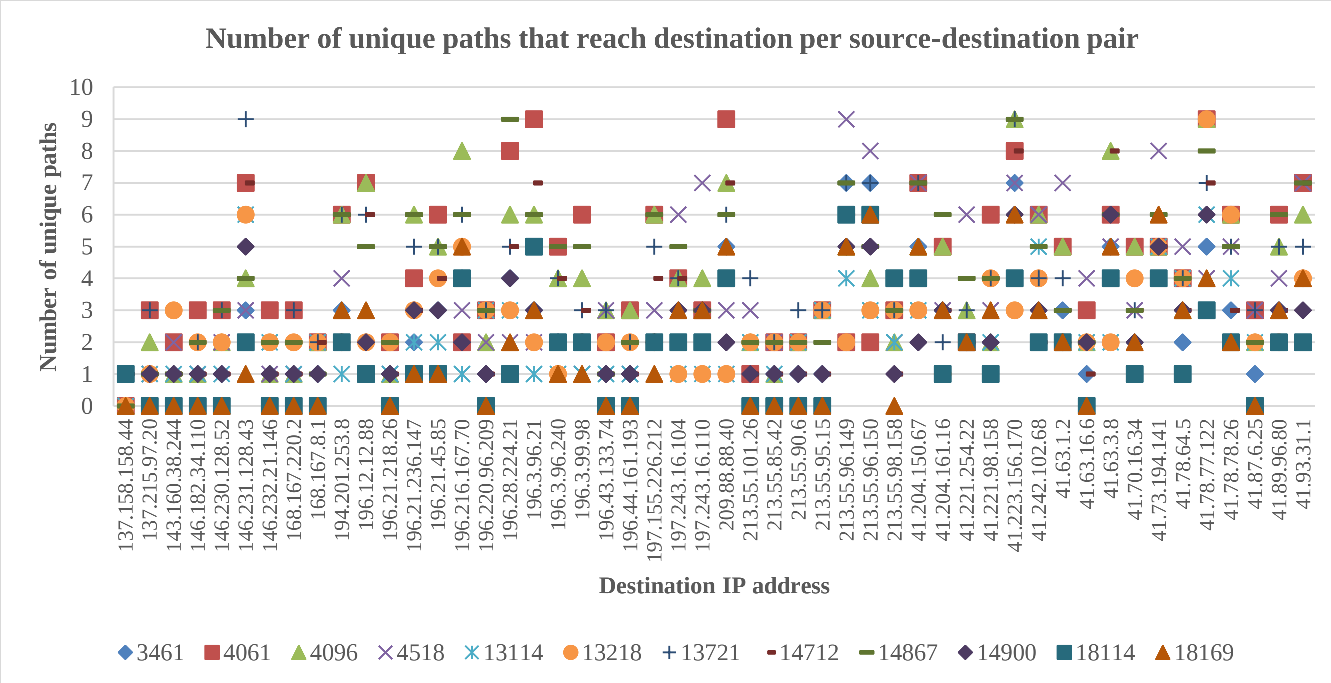

Figure 5. Graph of number of unique paths from a source to a destination that reached the destination

Of the source-destination pairs shown in Figure 5, eleven (2%) display nine unique paths. Since nine measurements were taken, this means that every probe reached the destination by a new path. In the example Figure 6 below, there are 9 unique paths from probe 14867 to destination 196.28.224.21. The source and destination are indicated in the Figure by a thick border.

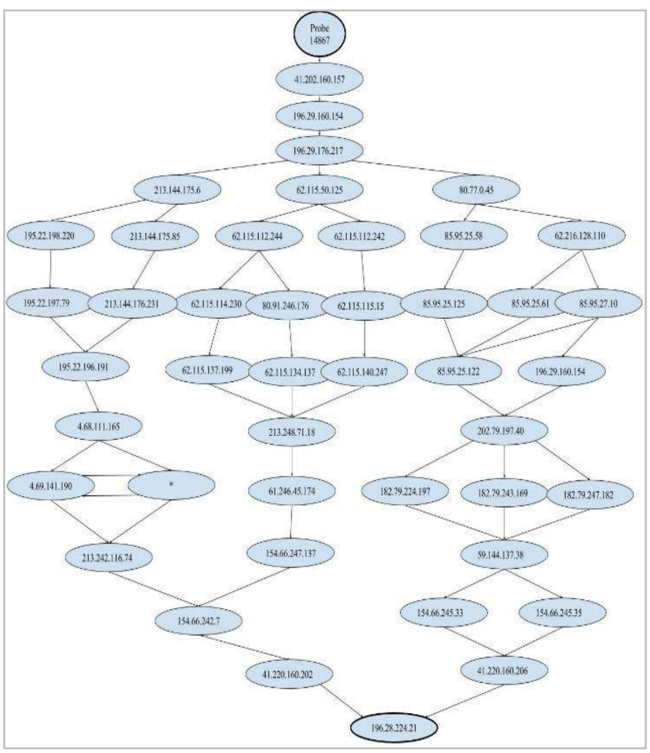

Figure 6. Example showing the path diversity from probe 14867 to destination 196.28.224.21

Path diversity can be expressed as the average number of unique paths between a source and a destination. As seen in Figure 7, for measurements that reached the destination, the path diversity is about 3. For all measurements, regardless of whether they reached the destination, the path diversity is almost 6.

Measurements that reach the destination have a lower diversity than that of all the measurements because it is a subset of all the measurements. There are fewer unique destination-reaching paths because there are fewer destination-reaching paths in general.

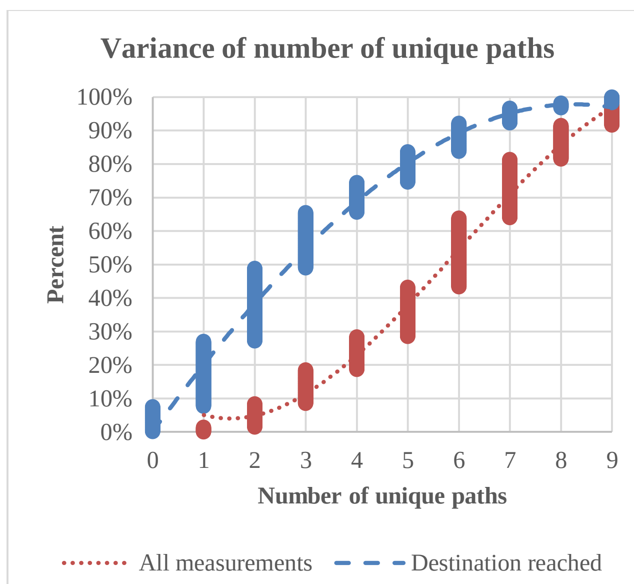

Figure 7. Cumulative Distribution Frequency graphs of number of unique paths for all measurements versus only measurements where the destination was reached

It is interesting to note that there is an increase in path discovery when new protocols are added. There are 600 source-destination pairs in total. Therefore there are 1800 measurements for each of the three protocols. When considering only ICMP-based measurements, there are 1168 unique paths. When TCP-based measurements are added to the dataset, there are 918 more unique paths. When UDP-based measurements are added to the combined datasets of ICMP and TCP-based measurements, 1266 more unique paths are discovered. This increasing trend in the number of unique paths indicates that different paths are found by using different protocols.

There are in fact 3352 unique paths in total. ICMP-based measurements have 1168 unique paths; TCP-based measurements have 1029 unique paths and UDP-based measurements have 1443 unique paths. This indicates that there is some overlap in unique paths between the three protocols. This overlap is displayed in the Venn diagram in Figure 8.

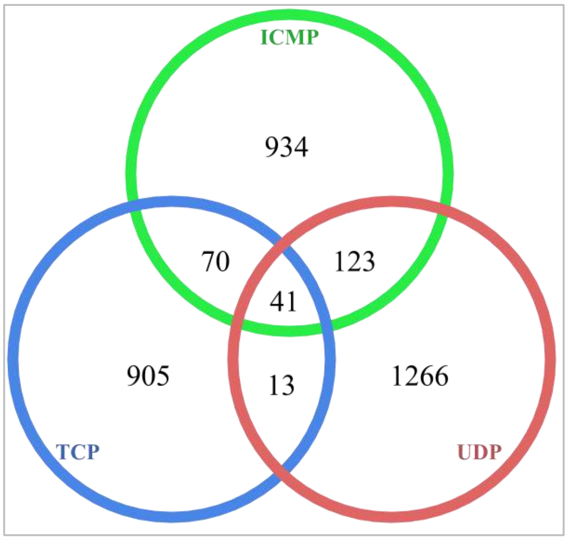

Figure 8. Venn diagram showing the overlap in the number of unique paths per protocol

Figure 9 shows the variance of the number of unique paths between a source and a destination for each protocol. This number can be either one, two or three. Trendlines have been added to each scatter plot to illustrate the variance better. As can be seen in Figure 9, the average path diversity for TCP-based measurements is the lowest at about 1.7 (59% out of 3 paths per source-destination pair). ICMP-based measurements have the next highest path diversity at almost 2 (64% out of 3 paths per pair). UDP-based measurements have the highest path diversity at about 2.4 (80% out of 3 paths per pair).

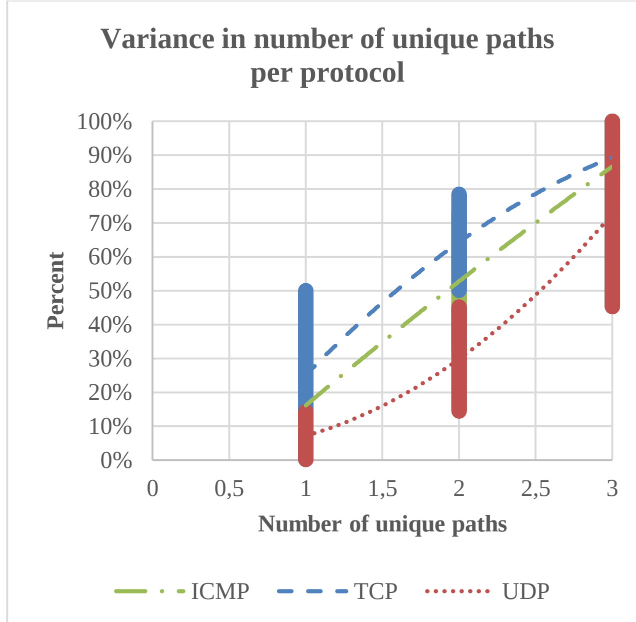

Figure 9. Cumulative Distribution Frequency graphs for number of unique paths per protocol

This adds to the argument that new paths are discovered by using different protocols as the most unique paths were found after adding the UDP-based measurements to the dataset. UDP-based measurements have the highest path diversity as well as the highest number of unique paths. Also, adding TCP-based measurements to the dataset only revealed 918 more unique paths. This is to be expected since TCP has the lowest path diversity out of the three protocols.

With respect to efficiency, on comparing the number of hops for full measurements and partial measurements, it was found that there was an average reduction of ten hops after modifying the first and/or last hop in the Traceroute measurement. There was therefore a reduction in probe packets.

4.3 Round Trip Times (RTTs)

Latencies were calculated as the RTT to the last hop. As shown in the Cumulative Distribution Frequency graphs in Figure 10, the average latency for each protocol is roughly the same at about 264 ms. However, as seen in Figure 10, there are more low latencies for TCP-based measurements than for UDP and ICMP-based measurements. About 40% of the latencies for TCP-based measurements are less than 100 ms, whereas only about 25% of the ICMP and UDP-based measurements are less than 100 ms.

The maximum latency for ICMP, TCP and UDP-based measurements is 2602.405 ms, 3972.355 ms and 2414.569 ms respectively. The graph is cut off at 1000 ms, though, to show the difference at the beginning more clearly. There is clearly a convergence towards 100% in the y-axis at the 1000 ms mark.

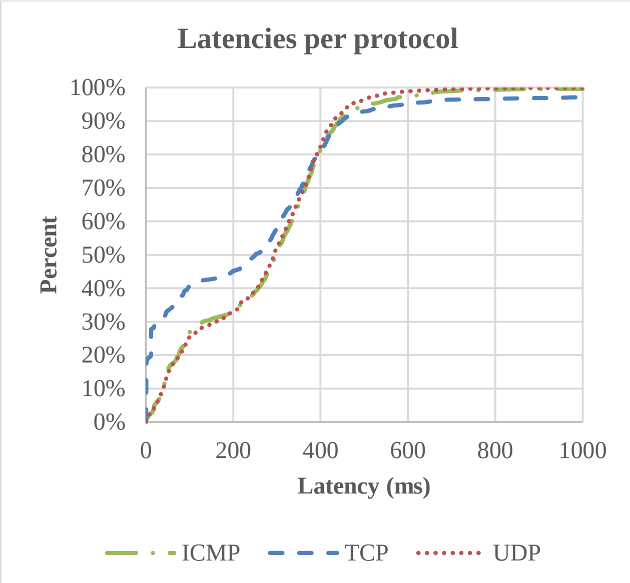

Figure 10. Cumulative Distribution Frequency graph showing distribution of latencies per protocol for full measurements

When only Traceroute measurements that reach the destination are considered, the average latency is about 293 ms per protocol and the cumulative frequency graphs generally follow the same curve for all three protocols. This indicates that, when considering only measurements that reach the destination, using a different protocol to determine the latency does not produce significantly different results.

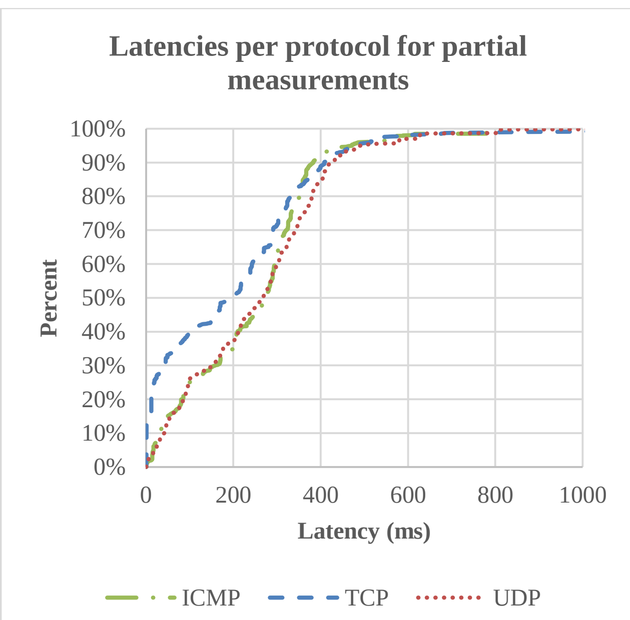

Figure 11. Cumulative Distribution Frequency graphs showing distribution of latencies for partial measurements where first and/or last hops were altered

Measurements where the first and/or last hop were edited have an average latency of 233 ms as shown in Figure 11. When comparing Figures 10 and 11, it can be seen that the curves follow a similar trend. This indicates that the alterations in the Traceroute measurements did not have a significant effect on the results.

4.4 Statistical Tests

An ANOVA test was conducted to deduce whether the Round Trip Times varied for a change in the Paris value. The data was separated by protocol so that only one variable (Paris value) was considered. The p-values from the test are displayed in Table 3. Since $p > 0.05$ for all protocols, the null hypothesis was rejected and we concluded that there was no significant difference at the 5% significance level. Because of this, the Paris value was removed as an independent variable for measurements in the experiment.

Table 3. P-values from ANOVA test to check for significant difference between Paris values

Protocol	P-value
ICMP	0.63
TCP	0.5
UDP	0.4

Another ANOVA test was run to see whether there was a significant difference in Round Trip Times for a change in protocol. The test showed that $p = 0.0005$. Since $p < 0.05$, the null hypothesis was not rejected and we concluded that there was a significant difference at the 5% significance level. Therefore, all three protocols were used.

5. ANALYSIS/DISCUSSION

5.1 Accuracy

The use of multiple protocols allowed more destinations to be reached. Measurements configured with UDP only reached target destinations 26% of the time, whereas measurements configured with TCP reached target destinations 63% of the time. Measurements configured with TCP also managed to reach all fifty destinations whereas ICMP and UDP measurements reached only 70% and 56% of the destinations respectively. Therefore, TCP measurements provided the most accurate set of data in terms of reachability.

Running multiple Traceroute measurements and using all three protocols led to the discovery of more paths. Even though TCP measurements reached all the destinations, they did not discover all the paths to those destinations. By using all three protocols, more paths were discovered. In fact, UDP measurements discovered the most unique paths out of the three protocols. This led to a more complete view of the topology being discovered. The accuracy was therefore increased by using all three protocols because a more complete topology gives more information on where traffic might be routed.

5.2 Efficiency

Finding overlapping paths and conducting experiments where those overlaps were taken into account did not cause a significant change in the data with respect to destinations reached and latencies. Therefore the partial measurements were just as accurate as the full measurements.

There was a reduction in the total number of probe packets of, on average, ten packets for the partial measurements. This indicates a general increase in efficiency.

6. CONCLUSIONS

Through the use of multiple protocols, multiple vantage points, a broad set of destination IP addresses, and by running the experiment multiple times, a fairly accurate depiction of the topology was discovered. Not all destinations were reached for each source-destination pair but all destinations were reached by at least one probe. Multiple paths were found between probes and destinations which increased the completeness of the topology discovered.

By conducting even more measurements at different times of day, it is likely that more diversity in paths will be found. It is also likely that more destinations will be reached by more probes.

Better algorithms could be found for finding overlapping paths and for increasing the efficiency of the measurements.

Future work should also aim to streamline the whole system to allow for new data to be collected, stored, aggregated and displayed in a visualisation at the click of a button. This would help researchers, network managers and other interested parties in planning new routing policies based on an accurate and up-to-date set of data.

7. ACKNOWLEDGEMENTS

Thank you to team members Chantal Yang and Robert Passmore for support during this project. Thank you also to supervisors Josiah Chavula and Assoc. Prof. Hussein Suleman for support and guidance in writing this paper and designing experiments. Thank you to Assoc. Prof. Michelle Kuttel for being second reader and for being willing to assist anyone at any time. Lastly, thank you to all those who proofread and offered advice and support in writing this paper.

8. REFERENCES

[1] Aben, E. Measuring countries and IXPs with RIPE Atlas, 2015. Retrieved 3 August 2015 from Ripe Network Coordination Centre: https://labs.ripe.net/Members/emileaben/measuring-ixps-with-ripe-atlas.

[2] Augustin, B., Cuvellier, X., Orgogozo, B., Viger, F., Friedman, T., Latapy, M., Magnien, C., and Teixeira, R. Avoiding Traceroute anomalies with Paris Traceroute. in 6*th ACM SIGCOMM conference on Internet measurement,* (New York, USA, 2006), ACM Press, 153-158.

[3] Chavula, J., Feamster, N., Bagula, A., and Suleman, H. Quantifying the effects of circuitous routes on the latency of intra-Africa Internet traffic: a study of research and education networks. in *6th International Conference, AFRICOMM,* (Kampala, Uganda, 2014), Springer International Publishing, 64-73.

[4] Cheswick, B., Burch, H., and Branigan, S. Mapping and Visualising the Internet. in *USENIX Annual Technical Conference,* (Berkeley, USA, 2000), USENIX Association, 1-1.

[5] Claffy, K., Hyun, Y., Keys, K., Fomenkov, M., and Krioukov, D. Internet mapping: from art to science. in *Conference for Homeland Security, 2009. CATCH'09. Cybersecurity Applications & Technology,* (Washington, DC, USA, 2009), IEEE, 205-211.

[6] Donnet, B., and Friedman, T. Internet topology discovery: a survey. *Communications Surveys & Tutorials, IEEE, 9* (4), 56-69.

[7] Fryer, T. Research and education networks around the world and their use. in Segunda Conferencia de Directores de Tecnologia de Informacion, (Lima, Peru, 2012), 1-13, http://hdl.handle.net/10786/847

[8] Gilmore, J., Huysamen, N., and Krzesinski, A. Mapping the African Internet. in *2nd IFIP International Symposium on Wireless Communications and Information Technology in Developing Countries,* (Pretoria, South Africa, 2007)

[9] Gupta, A., Calder, M., Feamster, N., Chetty, M., Calandro, E., and Katz-Bassett, E. Peering at the Internet's frontier: A first look at ISP interconnectivity in Africa. in *15th International Conference, PAM,* (Los Angeles, CA, USA, 2014), Springer International Publishing, 204-213.

[10] Harper, K.C., Chen, K., and Yen, D.C. Distance learning, virtual classrooms, and teaching pedagogy in the Internet environment. *Technology in Society, 26* (4), 585-598.

[11] Hyun, Y. Archipelago measurement infrastructure, 2015. Retrieved April 24, 2015 from Centre for Applied Internet Data Analysis: http://www.caida.org/projects/ark/.

[12] Hyun, Y. CAIDA Monitors: The Archipelago Measurement Infrastructure, 2015. Retrieved April 20, 2015 from Centre for Applied Internet Data Analysis: http://www.caida.org/data/monitors/monitor-map-ark.xml.

[13] Luckie, M., Hyun, Y., and Huffaker, B. Traceroute probe method and forward IP path inference. in *8th ACM SIGCOMM conference on Internet measurement,* (New York, USA, 2008), ACM Press, 311-324.

[14] Mao, Z.M., Rexford, J., Wang, J., and Katz, R.H. Towards an accurate AS-level Traceroute tool. in *2003 conference on Applications, technologies, architectures, and protocols for computer communications,* (New York, USA, 2003), ACM Press, 365-378.

[15] Motamedi, R., Rejaie, R., and Willinger, W. A Survey of Techniques for Internet Topology Discovery. *Communications Surveys & Tutorials, IEEE, 17* (2), 1044-1065.

[16] Pansiot, J., Mérindol, P., Donnet, B., and Bonaventure, O. Extracting intra-domain topology from mrinfo probing. in *11th International Conference, PAM,* (Zurich, Switzerland, 2010), Springer Berlin Heidelberg, 81-90.

[17] Ripe Atlas, 2010. Retrieved April 24, 2015 from Ripe Network Coordination Centre: https://atlas.ripe.net.

[18] Shavitt, Y., and Shir, E. DIMES: Let the Internet measure itself. *ACM SIGCOMM Computer Communication Review 35* (5), 71-74.

[19] Shavitt, Y., and Weinsberg, U. Quantifying the importance of vantage points distribution in internet topology measurements. in *INFOCOM 2009, IEEE,* (Rio de Janeiro, 2009), IEEE, 792-800.

[20] USC-NSL. ripe-atlas, 2015. Retrieved 7 September 2015 from Github: https://github.com/USC-NSL/ripe-atlas.