

ANALISA PENCARIAN JARAK ANTAR HOST MENGGUNAKAN PENDEKATAN KOORDINAT DALAM JARINGAN P2P UNTUK MENGEMBANGKAN METODE OPTIMALISASI JARINGAN

David Kurniawan Santoso A11.2006.03098, Teknik Informatika, Universitas Dian Nuswantoro

Abstrak - Berawal dari peningkatan kebutuhan manusia akan informasi dan hiburan melalui media digital, dalam hal ini internet sebagai perantara, membuat kebutuhan dalam penggunaan jaringan komputer semakin luas dan banyak. Untuk mendapatkan akses yang cepat, pada umumnya pengguna akan memperbesar bandwidth data, tetapi kita mencari solusi lain, penulis mencoba mengusulkan untuk mengoptimalkan proses routing dengan memperhatikan jarak host akhir dan awal dalam internet. Menggunakan mekanisme berbasis koordinat dalam arsitektur peer-to-peer untuk memprediksi jarak host dalam jaringan internet. Hasil yang diharapkan yaitu dapat lebih mengoptimalkan bandwidth yang sudah ada dengan pemilihan host yang terdekat. Kita mempelajari dua mekanisme. Yang pertama adalah skema, yang disebut heuristik Triangulasi, yang didasarkan pada relatif koordinat jarak dari host dari beberapa node jaringan khusus. Dan mekanisme kedua, yang disebut Global Network Positioning (GNP), yang didasarkan pada koordinat mutlak/absolut dihitung dari pemodelan internet sebagai geometris ruang euclidean. Sejak host akhir mempertahankan koordinat mereka sendiri, pendekatan ini memungkinkan host akhir untuk menghitung jarak antar-host mereka segera setelah mereka menemukan satu sama lain. Selain itu koordinat sangat efisien dalam menyimpulkan jarak antar-host, membuat pendekatan ini sangat bisa diukur. Dengan melakukan percobaan menggunakan data jarak internet yang telah diukur, kami menunjukkan bahwa kedua skema berbasis koordinat lebih akurat dibandingkan dengan yang sudah ada yaitu Pendekatan IDMaps.

I. PENDAHULUAN

Meningkatnya jumlah manusia di dunia, mengakibatkan kebutuhan akan segala sesuatu juga makin bertambah. Oleh karena itu dibutuhkan sebuah pengetahuan baru ataupun peningkatan kualitas dan cara-cara baru. Kini hampir setiap manusia di dunia pasti membutuhkan informasi, bisa informasi akan ilmu pengetahuan, hiburan, ekonomi,

olahraga, dan masih banyak lagi. Sebelum menuju era sekarang yang serba digital dan instan, informasi diperoleh dari media cetak, media digital fisik. Media cetak dapat berupa buku, majalah, koran, tabloid, makalah, novel. Sedangkan media digital fisik dapat berupa kaset pita (suara atau suara dan gambar), kepingan CD (Suara atau suara dan gambar).

Dengan bertambahnya sumber informasi dan pencari informasi di dalam jaringan, membuat kepadatan lalu lintas data bertambah dengan pesatnya, pada umumnya untuk mendapatkan kecepatan akses terhadap informasi melalui internet, kita cenderung untuk memperbesar bandwidth, memperbesar bandwidth memang dapat mempercepat akses selama bandwidth sumber lebih besar dibanding bandwidth penerima, tetapi memperbesar bandwidth sama saja dengan memperbesar biaya, oleh maka dari itu perlu cara lain yang lebih murah dan mudah penerapannya. Salah satunya adalah dengan optimalisasi jarak antar perangkat, hal ini sangat berpengaruh terutama di sistem jaringan P2P. Semakin dekat jarak, berarti data akan melewati lebih sedikit titik di internet, menghasilkan waktu tunggu yang lebih singkat juga, ini juga dapat mengurangi tingkat error dalam pengiriman data. Penyedia layanan

Untuk mengetahui sebuah jarak, kita dapat mengukurnya secara nyata, tetapi ini memakan biaya dan waktu. Lagipula, titik/node yang dilalui bisa berubah setiap waktunya. Untuk itu cara yang lebih efisien diperlukan, ada aplikasi bernama IDMaps yang mempunyai infrastruktur khusus untuk memprediksi jarak, yang mana apabila semakin banyak titik yang berfungsi khusus tersebut, maka akan lebih akurat. Kemudian ada mekanisme berbasis titik koordinat yang dikomputasikan kemudian dimodelkan sebagai ruang geometri internet. Untuk memprediksikan digunakan pendekatan Triangulated heuristic dan Global Network Positioning (GNP). Di sini pendekatan GNP dianggap paling tepat.

II. GLOBAL NETWORK POSITIONING

Adalah mekanisme berbasis koordinat dalam arsitektur jaringan peer-to-peer yang memprediksi jarak jaringan internet (yaitu round-trip propagasi dan delay transmisi). Mekanisme ini didasarkan pada koordinat absolut dihitung dari pemodelan internet sebagai ruang geometris. Sejak akhir host mempertahankan koordinat mereka sendiri, pendekatan ini memungkinkan end host untuk menghitung jarak antar-host mereka segera setelah mereka menemukan satu sama lain. Selain itu, koordinat sangat efisien dalam meringkas jarak antar-host, membuat pendekatan yang sangat scalable

III. METODE PENGUMPULAN DATA

A. Pengumpulan Data

Pada penelitian ini digunakan akses ke 19 host yang disebut Probes(satelit) tersebar di seluruh dunia. 12 berada di Amerika utara, 5 di asia pasifik, dan 2 di eropa. Untuk melengkapi probes tersebut, kita telah mengkompilasi sejumlah alamat IP untuk merespon pesan ping ICMP. Kemudian disebut IP target. Koordinat ditentukan dari Round Trip Time (RTT), RTT dari host pengirim ke penerima dan RTT dari penerima ke pengirim. RTT dipilih yang paling minimum dari sekitar 10 kali percobaan.

Untuk mengumpulkan kesatuan data, kita mengukur jarak diantara probes dan jarak dari setiap probe ke sebuah target yang ditentukan. Untuk mengukur jarak diantara 2 host, kita mengirim 220 paket 84byte ICMP ping dengan jeda waktu satu detik dan mengambil nilai minimum round-trip waktu perkiraan dari semua balasan sebagai jarak. Data mentah ini kemudian diproses untuk mempertahankan hanya target yang bisa dijangkau dari semua probe. Sejalan dengan itu, ada bias terhadap target yang tidak selalu-hidup (misalnya modem host) atau tidak memiliki konektivitas global dalam target akhir yang telah ditetapkan.

Kami telah mengumpulkan dua set data. Set pertama, dikumpulkan selama periode dua-hari di minggu terakhir Mei 2001, didasarkan pada set target yang berisi 2.000 alamat IP "ping-able" yang diperoleh pada waktu sebelumnya. Alamat IP tersebut dipilih melalui penyelidikan di ruang alamat IP bahwa setiap alamat IP yang valid memiliki kesempatan yang sama untuk terpilih. Setelah diolah lebih lanjut, kita memiliki 869 target yang dapat dicapai darisemua probe. Hasil relatif rendah dimungkinkan beberapa target tidak tersambung dengan internet selama pengukuran kami lakukan, dan sebagian karena kemungkinan bahwa beberapa target yang tidak terjangkau secara global karena kegagalan parsial dari Internet. Menggunakan NetGeo alat dari CAIDA, kami telah menemukan bahwa 869 target menjangkau 44 negara yang berbeda. 467 target berada di Amerika Serikat, dan masing-masing negara memberikan kontribusi yang kurang dari 40 target. Singkatnya, 506 target berada di amerika utara, 30 target berada di Amerika Selatan, 138 target di Eropa, 94 target berada di Asia, 24 target yang di Oceania, 12 target di Afrika, dan 65 target memiliki lokasi yang tidak diketahui.

Set data global ini memungkinkan kita untuk mengevaluasi penerapan secara global untuk mekanisme prediksi jarak yang berbeda.

Set kedua dikumpulkan dalam periode selama 8 jam di minggu pertama Juni 2001, didasarkan pada satu set 164 target server web institusi yang terhubung ke jaringan backbone Abilene. Setelah post-processing, kita memperoleh 127 target yang bisa dijangkau dari semua probe. Sebagian besar ini target berada di universitas - universitas di Amerika Serikat. Perhatikan bahwa 10 dari 19 probe kami juga terhubung ke Abilene. Abilene ini merupakan kumpulan data yang memungkinkan kita untuk memeriksa kinerja dalam lingkungan yang lebih homogen.

B. Langkah langkah

Ketiga mekanisme prediksi jarak yang dipertimbangkan (IDMaps, Triangulated Heuristic dan GNP) memerlukan penggunaan beberapa node infrastruktur khusus (i-node). Untuk melakukan percobaan menggunakan satu set data, pertama-tama kita pilih subset dari 19 probe untuk digunakan sebagai i-node, dan menggunakan probe dan target yang tersisa sebagai host biasa. Dengan cara ini, kita dapat mengevaluasi kinerja mekanisme dengan langsung membandingkan jarak prediksi dan jarak telah diukur dari probe yang tersisa dengan target. Karena pemilihan i-node yang berpotensi dapat mempengaruhi keakurasian prediksi yang dihasilkan.

Namun ada masalah yang penting dan harus diperhatikan. Misalkan kita ingin membandingkan GNP ke IDMaps. kita bisa memilih kriteria seleksi untuk memilih N i-node dan melakukan satu percobaan menggunakan GNP dan satu menggunakan IDMaps. Sayangnya, ketika kita membandingkan hasil, sulit untuk menyimpulkan apakah perbedaannya adalah karena perbedaan yang melekat dalam mekanisme ini, atau hanya karena fakta bahwa set tertentu dari i-node bekerja lebih baik dengan salah satu mekanisme. Untuk meningkatkan keyakinan dalam hasil kami, kami menggunakan teknik yang mirip dengan k-fold validasi di pembelajaran mesin. Alih-alih memilih N i-node berdasarkan kriteria, kita memilih N + 1 i-node. kemudian dengan menghilangkan salah satu dari N + 1 i-node pada suatu waktu, kita bisa menghasilkan N + 1 set yang berbeda dari N i-node yang cukup dekat untuk memuaskan kriteria untuk N. Kami kemudian membandingkan mekanisme yang berbeda dengan menggunakan hasil keseluruhan dari semua N + 1 set dari N i-node.

Untuk mengatasi permasalahan minimisasi global yang multi-dimensi dalam menghitung koordinat GNP, kita menggunakan Simplex Metode Downhill. Metode ini sangat kuat dan cukup efisien. Untuk memastikan solusi berkualitas tinggi, kita ulangi prosedur minimisasi untuk 300 iterasi ketika mengolah koordinat Landmark, dan 30 iterasi ketika mengolah koordinat host biasa itu. Dalam prakteknya, 3 iterasi sudah cukup untuk mendapatkan perkiraan yang cukup kuat.

Secara intuisi, kami ingin i-node yang akan terdistribusi dengan baik sehingga informasi yang bermanfaat yang

mereka berikan dimaksimumkan. berdasarkan intuisi ini, kami mengusulkan tiga kriteria untuk memilih strawman N i-node dari 19 probe. Kriteria pertama, yang disebut pemisahan maksimum, adalah memilih N probe yang punya jarak maksimal antar probe. Kriteria kedua, yang disebut N-median, adalah memilih N probe yang punya total jarak minimal dari setiap probe yang tidak terpilih ke probe terpilih paling dekat. Kriteria ketiga, yang disebut N-cluster median, adalah untuk membentuk N kluster probe dan kemudian memilih median dari masing-masing cluster sebagai i-node. N cluster dibentuk oleh Penggabungan iteratif dua cluster terdekat, dimulai dengan 19 Probe cluster, sampai tersisa dengan N cluster. Selain itu, untuk mengamati bagaimana masing-masing mekanisme prediksi bereaksi untuk berbagai pilihan i-node secara asal, kita akan juga menggunakan kombinasi acak dari i-node dalam penelitian ini.

C. Pengujian

Untuk mengukur seberapa akurat hasil dari prediksi jarak dengan jarak sesungguhnya, kita menggunakan sebuah metric yang disebut directional relative error yang didefinisikan

$$\frac{\text{predicted distance} - \text{measured distance}}{\min(\text{measured distance}, \text{predicted distance})}$$

Dengan demikian, nilai nol menyiratkan prediksi yang sempurna, nilai satu menyiratkan jarak diprediksi lebih besar dari dua faktor, dan nilai negatif berarti jarak diprediksi adalah lebih kecil dari dua faktor. Dibandingkan dengan simple percentage error, metrik ini dapat terjaga dari aturan "selalu memprediksi nol". Ketika mempertimbangkan akurasi prediksi umum, kita juga akan menggunakan relative error metric, untuk memastikan nilai dari directional relative error.

Untuk mengukur efektivitas dari penggunaan jarak prediksi untuk tipe aplikasi pilihan-server, kami menggunakan metrik yang disebut rank accuracy(peringkat akurasi). Idenya adalah bahwa, setelah setiap percobaan, kami memiliki jarak diprediksi dan jarak terukur untuk jalur antara probe non-i-node dan target. Kami kemudian mengurutkan basis jalur pada jarak prediksi untuk menghasilkan daftar peringkat prediksi. Peringkat keakuratan kemudian didefinisikan sebagai persentase jalan yang dipilih dengan benar ketika kita menggunakan daftar peringkat prediksi untuk memilih beberapa angka dari jalan terpendek. Jika peringkat prediksi sempurna, maka akurasi rank adalah 100% terlepas dari jumlah jalur terpendek kita memilih. Perhatikan bahwa mekanisme prediksi berpotensi dapat sangat tidak akurat sehubungan dengan kesalahan directional relative error metric tetapi masih memiliki akurasi peringkat tinggi karena peringkat dari jalur masih dapat dipertahankan.

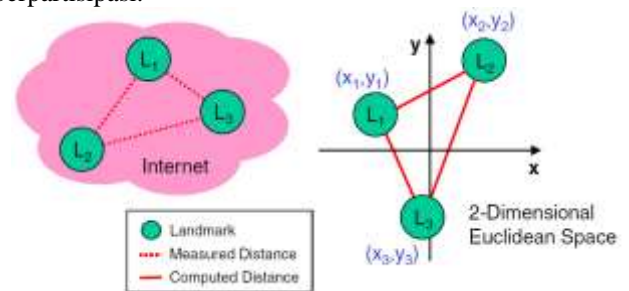
IV. HASIL DAN PEMBAHASAN

A. Pemodelan Internet Sebagai Ruang Geometri Terukur

Untuk dapat membuat sebuah koordinat host di internet yang dapat diukur, dilakukan dua langkah untuk membuat arsitekturnya, yaitu :

1. Landmark Operations

Langkah pertama diberi nama *Landmark operation*, beberapa kumpulan host yang tersebar ditentukan kemudian disebut Landmark, kemudian landmark ini akan menghitung koordinat mereka sendiri di ruang geometri bersangkutan. Koordinat landmark berfungsi sebagai acuan dan disebarluaskan kepada host mana pun yang ingin berpartisipasi.



Gambar 1 : Operasi Landmark

Kita ingin memodelkan internet sebagai ruang (*Space*) geometri dengan S , Menunjukkan koordinat host H di S sebagai C_H^S , jarak yang berfungsi di koordinat tersebut disebut F^S , dan jarak antara H_1 dan H_2 adalah $F^S(C_{H_1}^S, C_{H_2}^S)$ disederhanakan menjadi $d_{H_1 H_2}^S$.

Anggap ada sejumlah N landmark, L_1 sampai L_N . Landmark mengukur waktu di berputar di dalam landmark itu sendiri menggunakan pesan ping ICMP, dan mengambil nilai minimum dari beberapa pengukuran untuk setiap jalur untuk menentukan garis dasar dari jarak acuan $N \times N$ (acuan dianggap simetris sepanjang diagonal). Kita menunjukkan jarak antara host H_1 dan H_2 sebagai $d_{H_1 H_2}$. Menggunakan jarak terukur $d_{L_i L_j}$, $i > j$, sebuah host, mungkin berada di salah satu N landmark, menghitung koordinat landmark di S . Tujuannya untuk mencari kumpulan koordinat $C_{L_1}^S, \dots, C_{L_N}^S$, untuk landmark N , supaya error antara jarak terukur dengan jarak yg dihitung dalam S diminimalkan. Kita mencari untuk meminimalkan fungsi objek f_{obj1} :

$$f_{obj1}(C_{L_1}^S, \dots, C_{L_N}^S) = \sum_{L_i L_j \in \{L_1, \dots, L_N\} | i > j} \mathcal{E}(d_{L_i L_j}, d_{L_i L_j}^S)$$

Dimana \mathcal{E} adalah fungsi ukur error, yang bisa menjadi squared error

$$\mathcal{E}(d_{H_1 H_2}, d_{H_1 H_2}^S) = (d_{H_1 H_2} - d_{H_1 H_2}^S)^2$$

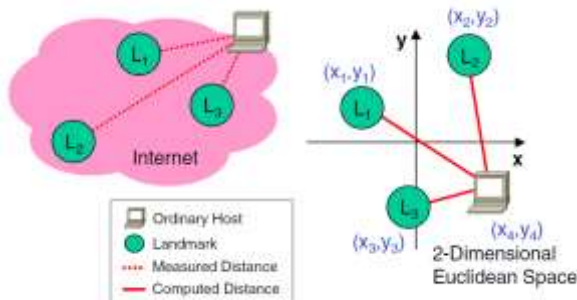
atau menggunakan ukur error yang lebih mutakhir. Cara mengukur error di fungsi objek sangatlah berpengaruh di keakuratan prediksi. Dengan formulasi ini, penghitungan koordinat dapat disebar sebagai hal umum untuk meminimalkan masalah secara global dan multi dimensi, yang diperkirakan dapat diselesaikan dengan banyak metode seperti metode Simplex Downhill. Gambar 4.1 mengilustrasikan operasi landmark dengan 3 landmark dalam ruang 2 dimensi euclidean. Catat bahwa ada solusi tak terbatas untuk koordinat landmark karena sedikitpun putaran dan atau tambahan penerjemahan set koordinat akan merubah jarak antar landmark. Tetapi karena koordinat landmark hanya digunakan sebagai kerangka kerja referensi dalam GNP, hanya lokasi kekerabatan yang penting. Ketika

dibutuhkan penghitungan ulang koordinat dibutuhkan seiring waktu, kami dapat pastikan koordinat tidak berubah drastis jika kita masukkan koordinat lama dibanding nilai acak ketika kita memulai meminimalkan masalah.

Sesaat setelah koordinat landmark, $C_{L1}^S, \dots, C_{LN}^S$ dihitung, kemudian disebarkan, bersama dengan pengenalan untuk penggunaan ruang geometri S dan jarak F^S ke host biasa yang ingin bergabung dalam GNP, kita mengabaikan mekanisme penyebaran (misal unicast, multicast, push, pull) dan spesifikasi protokol

2. Ordinary Host Arsitektur

Langkah ke dua untuk memperlengkapi landmark dalam ruang S , dibutuhkan peran aktif dari host yang tidak masuk dalam landmark. Diperlengkapi oleh koordinat dari landmark, setiap host biasa (yang tidak termasuk dalam landmark) dapat menghitung koordinat diri mereka sendiri, menyesuaikan dengan host-host di landmark.



Gambar 2 : Operasi Ordinary Host

Untuk melakukannya, host biasa H menghitung waktu perjalanan ke N landmark menggunakan pesan ping ICMP dan mengambil nilai terkecil dari beberapa kali pengukuran jalur sebagai jarak. Pada tahap ini, landmark menjadi pasif dan hanya menjawab pesan ping ICMP. Menggunakan hasil pengukuran host ke landmark d_{HLi} tersebut, host H dapat ditentukan koordinatnya C_H^S yang meminimalkan error antara jarak terukur dengan jarak komputasi host ke landmark. Kita mencari untuk meminimalkan fungsi objek f_{obj2} :

$$f_{obj2}(C_H^S) = \sum_{Li \in \{L1, \dots, LN\}} \mathcal{E}(d_{HLi}, d_{HLi}^S)$$

Dimana \mathcal{E} adalah fungsi ukur error sama seperti sebelumnya. Gambar 4.2 mengilustrasikan operasi ini untuk host biasa dengan 3 landmark.

Inilah mengapa jumlah N landmark harus lebih besar dari dimensi D dalam ruang S . Jika N tidak lebih besar dari D , koordinat landmark dipastikan bohong di dimensi $D - 1$. Konsekuensinya, sebuah titik di dimensi D dan gambaran melalui landmark hyperplane tidak dapat dibedakan melalui fungsi objek, membuat ke ambiguan koordinat host. Catat juga bahwa tidak ada jaminan bahwa koordinat host akan selalu unik. Menggunakan dimensi yang lebih kecil dari landmark adalah cara termudah untuk menghindarinya.

B. Hasil

Di bagian ini kami jabarkan hasil penelitian. Pertama kami menggunakan set i-nodes yang sama untuk setiap mekanisme pengujian, hasil digunakan untuk membandingkan keakuratan metode GNP, triangulated

heuristic dan IDMaps. Kemudian kita membandingkan efektifitas dengan semua mekanisme dengan tiga i-node pilihan.

a. Pembandingan Menggunakan Set Data Global

Menggunakan cara *Relative Error* dan 6 i-node serta 15 i-node. Untuk GNP, hasil terbaik diperoleh saat menggunakan 5 dan 7 dimensi. Untuk triangulated heuristic, heuristic batas atas (U) menghasilkan yang terbaik sejauh ini. Ingat bahwa U adalah jarak terdekat antara end host dalam satu i-node. Kedua metode berbasis koordinat menghasilkan jauh lebih baik dibanding IDMaps. Dengan GNP memperoleh hasil tertinggi secara *overall* di semua kasus. Dengan menggunakan 15 landmark, GNP dapat memprediksi hampir 90% jalur yang tersedia dengan relative error 0.5 atau lebih sedikit. Kita juga mencoba dengan 9 i node dan 12 i-node. Untuk rangkumannya dapat dilihat pada tabel 1:

Jumlah i-node	15	12	9	6
GNP	0.5/7D	0.58/7D	0.69/5D	0.74/5D
Triang / U	0.59	0.69	0.8	1.05
IDMaps	0.97	1.09	1.16	1.39

Tabel 1 : Ringkasan 90% jalur dengan relative error

Dari hasil tersebut terlihat jelas, semakin banyak jumlah i-nodes, menghasilkan hasil yang lebih bagus untuk ketiga mekanisme, dengan GNP menjadi yang paling akurat dibanding semuanya. Bagaimanapun, keakuratan IDMaps dan triangulated heuristic secara bertahap dapat menjadi lebih tinggi dari GNP jika jumlah i-nodes bertambah.

Kemudian kita memeringkat kemampuan untuk menemukan jalur terpendek dengan rank accuracy metric dengan 15 i-node digunakan. Kemampuan untuk memeringkat jalur terpendek dengan tepat diperlukan sekali karena masalah pemilihan server. *Overall*, GNP lebih akurat dalam memeringkat jalur terpendek. Secara khusus GNP lebih akurat ketika memprediksi 5% total jarak dibanding triangulated heuristic meskipun perbedaan error relatif mereka cenderung kecil. Sedangkan, walaupun IDMaps punya kinerja buruk dalam eror relatif, IDMaps dapat lebih bagus dalam memeringkat jarak terpendek dibanding triangulated heuristic.

b. Pembandingan Menggunakan Set Data Abilene

Sekarang kita meneliti menggunakan data host yang terhubung dengan abilene backbone. Kita hanya menggunakan 10 Abilene probe. Membandingkan 3 mekanisme dengan menggunakan enam dan sembilan i-node. enam i-node dipilih menggunakan kriteria N -cluster-median dengan validasi k -fold, tetapi untuk sembilan i-node, diperoleh dengan cara yang sangat mudah, cukup menghilangkan satu probe dari sepuluh probe terpilih. Menyediakan pilihan sepuluh kombinasi berbeda untuk sembilan i-node. Untuk GNP, performa terbaik diperoleh dengan menggunakan lima dan delapan dimensi, dan untuk

triangulated heuristic, lagi-lagi batas atas U menghasilkan akurasi yang lebih baik dibanding batas bawah atau rata-rata dari keduanya. Perhatikan jika dalam lingkungan yang sama di abilene, keakuratan ketiga mekanisme hampir tidak berubah dari 6 dan 9 i-node. Kami yakin jika hanya penambahan i-node dalam lingkungan yang homogen tidak banyak menambah informasi.

Dibandingkan dengan hasil sebelumnya berdasar set data global dengan sembilan i-node, saat 90% jalur yang tersedia terprediksi, tingkat relatif eror yang diperoleh GNP 0.56, triangulated heuristic 0.88, dan IDMaps 1.72. Terlihat hanya GNP yang mengalami peningkatan di dalam lingkungan abilene yang homogen. Kita meyakini bahwa semua jalur di abilene sangat pendek lebih pendek dari 70ms. Keunggulan GNP ada di prediksi jarak pendek.

Dari hasil mekanisme pemeringkatan di abilene ketika sembilan i-node digunakan, keunggulan GNP dalam prediksi jarak pendek sangat jelas. GNP dapat memprediksi jarak terdekat ketika sebanyak 7% dari jalur ditemukan. Sedangkan Triangulated heuristic hanya 4%, lebih rendah dari IDMaps yang 5%. Tetapi mekanisme triangulated mengungguli IDMaps ketika jalur yang ditemukan sebanyak 24%.

c. Pengaruh Peletakan Node

Meskipun triangulated heuristic sangat mudah, tetapi itu sangat kekurangan ketahanan karena akurasi sangat bergantung pada jumlah dan lokasi node dasar di jaringan.

Untuk mempelajari bagaimana pengaruh peletakan i-nodes secara sembarangan pada GNP, triangulated heuristic dan IDMaps, kita melakukan penelitian dengan dua puluh kombinasi acak dari enam i-nodes menggunakan set data global. Untuk setiap mekanisme dan setiap 20 kombinasi acak, kita menghitung nilai 90% eror relatifnya. Tabel 2 menunjukkan hasilnya. Dari tiga mekanisme, keakuratan GNP paling tinggi. Karena GNP tidak menggunakan topologi virtual, ini sangat kuat bahkan dalam peletakan i-node acak.

	Max	Min	Mean	Std Dev
GNP	0.94	0.65	0.7375	0.06906
Triang/U	1.37	0.66	0.8685	0.1686
IDMaps	1.84	1.0	1.287	0.2308

Table 2 : Ringkasan 90% jalur dengan relative error

d. Pemilihan Node

Jika sebelumnya pemilihan menggunakan metode N -cluster-median i-node, kali ini kita menggunakan 3 kriteria. Menggunakan set data global, kita menggunakan 3 kriteria dengan 6 dan 9 i-node (dengan validasi k -fold) dan menghitung 90% eror relatif untuk tiap set penelitian. Tidak lupa kita juga membandingkan dengan metode triangulated heuristic dan IDMaps. Tabel 3 menampilkan hasilnya.

$N = 6$	N -Cluster-median	N -median	Max sep.
GNP	0.74	0.78	1.04
Triang/U	1.05	1.08	4.64
Triang/L	1.85	1.53	1.93
Triang/(L+U)/2	1.53	1.31	3.3
IDMaps	1.39	1.43	5.57

$N = 9$	N -Cluster-median	N -median	Max sep.
GNP	0.68	0.7	0.83
Triang/U	0.8	0.77	1.19
Triang/L	2.06	2.0	2.11
Triang/(L+U)/2	1.43	1.38	1.69
IDMaps	1.16	1.09	1.74

Tabel 3 : 90% relative error sesuai kriteria pemilihan i-node

N -cluster-median dan N -median hampir mirip. Di bagian satunya kriteria pemisahan maximal bekerja dengan buruk karena kriteria ini cenderung untuk memilih probe yang hanya ada di Eropa dan Asia dan mereka tidak tersebar dengan baik. Perbandingan dengan tabel 4.2 menunjukkan bahwa kriteria N -cluster-median tidak optimal karena sudah ada kombinasi dari 6 node yang dapat menunjukkan rendahnya eror relatif senilai 0.65, 0.66, 1.0 secara berturut turut untuk GNP, triangulated heuristic dan IDMaps. Ingat bahwa batas bawah triangulated L tidak punya kekuatan untuk dibandingkan dengan batas atas triangulated U . Sejak filter maksimal digunakan di metrik L , menjadi lebih sensitif terhadap data yang bohong. Kenyataan bahwa U bekerja lebih baik secara tak langsung menyatakan bahwa routing jalur terpendek masih dapat diterima mendekati perkiraan utama dari kasus. Ini merupakan pengecualian. Ketika 6 i-node dipilih dari kriteria batas maksimal, metrik L bekerja lebih baik dari metrik U . Melihat dari set i-node, kita menemukan bahwa pengecualian itu untuk satu i-node di Kanada, semua i-node lainnya terletak di Asia dan Eropa. Ini sungguh menarik, karena sesungguhnya target utama berada di Amerika Utara, mereka berada diantara dimana i-node terbanyak berada. Lalu kita menemukan konfigurasi yang tepat dimana metrik L menjadi sangat akurat.

Kita juga melihat peringkat keakuratan dari triangulated heuristic. Untuk 6 i-node tidak ada kejutan, perbedaan peringkat akurasi dari U , L , dan $(L + U)/2$ metric sama dengan perbedaan mereka di eror relatif. Untuk 9 i-node, L , dan $(L + U)/2$ metric mempunyai peringkat yang lebih tinggi sebanyak 5-12% dibanding metrik U yang hanya 1%. Perbedaan peringkat akurasi sejalan dengan perbedaan di

eror relatif. Pembelajaran lebih lanjut perlu dilakukan untuk anomali ini.

e. Sumber Ketidakakuratan

Sejauh ini kami hanya menunjukkan perbedaan di akurasi dari 3 skema prediksi jarak, tetapi ketidakakuratan dan perbedaan tidaklah terlalu jelas. Disini kami jabarkan beberapa sumber ketidak akuratan :

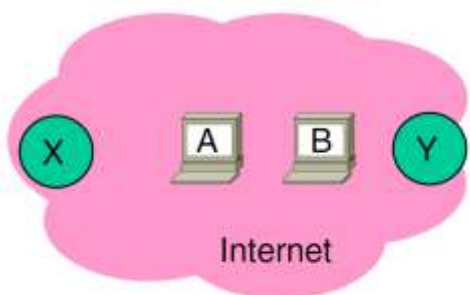
1. Routing yang tidak efektif

Karena dari awal prediksi jarak bergantung pada sudut terpendek dari jalur routing di internet, kami percaya bahwa yang menyebabkan ketidakakuratan adalah cara routing yang tidak efisien dari BGP (*Border Gateway Protocol policy routing and hop count based routing*). Untuk menaksir tingkat ketidakefisien routing di set data global kita melakukan tes triangular inequality sama dengan di [10]. Untuk semua jalur loop tertutup (a,b) , (b,c) , dan (a,c) yang kita ukur. Kita menghitung semua rasio $(a,b) / ((a,b) + (b,c))$.

Kita menemukan bahwa 7% dari ratio bernilai lebih dari 1, yang berarti konsisten dengan penemuan sebelumnya. Untuk mengukur dampaknya dengan prediksi akurasi, dilakukan percobaan sebagai berikut. Untuk semua target t in set data global, t dihilangkan dengan pertimbangan jika t didalam $\{a,b,c\}$ dan $(a,c)/((a,b) + (b,c)) > 1.5$. Setelah menerapkan filter ini, tersisa 392 target. Dilakukan lagi percobaan dengan 15 i-node dan menemukan bahwa ke tiga skema prediksi jarak mengalami peningkatan. Untuk GNP 90% eror relatif bertambah dari 0.5 menjadi 0.33; untuk triangulated heuristic/u bertambah dari 0.59 menjadi 0.42; dan untuk IDMaps bertambah dari 0.97 menjadi 0.89.

2. Prediksi jarak pendek

Perbedaan utama dari tiga skema berada di kemampuan untuk memprediksi jarak pendek. Sesuai yang kita tunjukkan, GNP merupakan yang paling akurat di kategori ini dan IDMaps merupakan yang paling kurang akurat dan cenderung over prediksi. Perbedaannya sangat mudah untuk dijelaskan.



Gambar 3 : Memperkirakan jarak pendek

Menurut contoh pada gambar 3 X dan Y adalah i-node dan A dan B adalah 2 host akhir yang sangat dekat. IDMaps memberi prediksi $(A,X) + (B,Y) + (X,Y)$, sedangkan triangulated heuristic memberi prediksi $(A,Y) + (B,Y)$. Secara kontras dengan model satu dimensi, GNP dapat dengan sempurna memprediksi jarak A dan B .

GNP bekerja lebih baik karena memanfaatkan hubungan antara posisi landmark dengan host akhir daripada bergantung pada lokasi pasti topologi dari i-node, meskipun itu akurat dan kuat.

4.2.6 Kerangka Kerja Global Network Positioning

1. Fungsi mengukur eror

Ingat kembali ketika mengukur koordinat GNP, fungsi ukur eror digunakan di fungsi objektif. Digunakan untuk memprediksi kekuatan koordinat. Kita menggunakan metode *Normalized Error*. Karena *Squared error* kurang tepat diterapkan karena satu unit eror untuk jarak sangat dekat untuk sama banyaknya dengan satu unit eror di jarak yang sangat panjang. Maka kita membandingkan metode normalisasi dan logaritma transform eror.

Metode Normalization error measure :

$$\epsilon(d_{H1H2}, d_{H1H2}^S) = \left(\frac{d_{H1H2} - d_{H1H2}^S}{d_{H1H2}} \right)^2 \dots \dots \dots (4-4)$$

Metode Logarithmic Transformed error measure :

$$\epsilon(d_{H1H2}, d_{H1H2}^S) = (\log(d_{H1H2}) - \log(d_{H1H2}^S))^2 \dots \dots (4-5)$$

Hasil dari keduanya relatif sama, karena sama-sama bentuk dari *relative error measure*.

2. Memilih Ruang Geometri

Meskipun dalam percobaan sebelumnya kita selalu melaporkan hasil dengan model ruang euclidean, selain itu kita juga mencoba dengan *spherical surface* dan *cylindrical surface* sebagai model yang potensial. Model spherical surface masuk akal karena bumi bulat, dan tidak adanya komunikasi yang melalui dua kutub model cylindrical surface juga bisa menjadi perkiraan yang bagus.

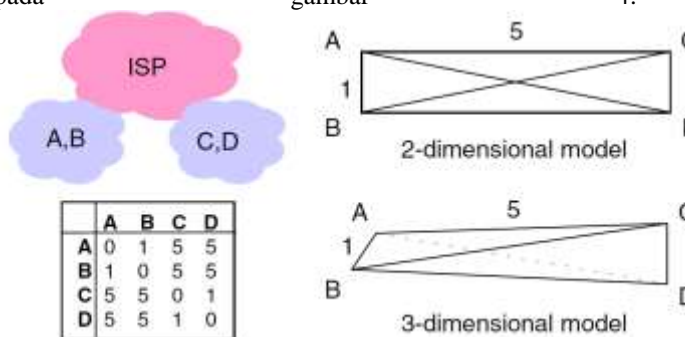
Kerangka kerja GNP sangat fleksibel untuk mengakomodasi dua model tersebut, perbedaannya hanya terdapat pada fungsi jarak. Dengan set data global dan 6 landmark terpilih dengan kriteria N -cluster-medians, kita mencoba melakukan percobaan untuk memeriksa kemampuan dari spherical dan cylindrical surface dengan berbagai ukuran.

Untuk spherical surface kita menentukan luasnya, untuk cylindrical surface kita menentukan keliling dan tingginya setengah dari keliling, ini membuat kinerja kedua model bertambah seiring dengan penambahan luas permukaan ruang dan dalam batas pendekatan kinerja model ruang dua dimensi euclidean. Ini adalah konsekuensi dari kenyataan bahwa kita tidak memiliki probes di Asia Tengah atau Afrika dan hanya sedikit target di wilayah tersebut, sehingga tidak membantu.

Fokus pada model ruang euclidean, kita berpikir berapa banyak dimensi yang harus digunakan di GNP. Untuk menjawabnya dilakukan percobaan dengan set data global menggunakan 6,9,12 dan 15 landmark terpilih dengan

kriteria N -cluster-median (dengan validasi k -vold) dengan berbagai jumlah dimensi. Pada umumnya bertambahnya dimensi meningkatkan akurasi GNP, tetapi penambahan berkurang dengan setiap dimensi berturut-turut. Untuk menggambarkan efek ini, mempertimbangkan fungsi cumulative probability distribution dari error relative dalam dua dimensi D dan $D+1$. Antara 70 dan 90% performa dimensi $D+1$ tidak bertambah banyak dari dimensi D , atau rata-rata penambahan kurang dari 0.1%, hasil digabung di dimensi D . Menggunakan kriteria ini untuk 6,9,12 dan 15 landmark, hasilnya bertemu masing-masing di dimensi 5,5,7 dan 7.

Seharusnya menambah dimensi meningkatkan fleksibilitas model dan membuat perhitungan koordinat lebih akurat. Untuk mengilustrasikannya pertimbangkan situasi pada gambar 4.



Gambar 4 : Keuntungan dimensi ekstra

Dimana ada 4 host, A, B, C dan D dengan A di jaringan yang sama dengan B besar dan C di dalam jaringan yang sama dengan D. Dalam ruang 2 dimensi jarak tidak dapat dimodelkan dengan sempurna. Satu perkiraan yang mungkin adalah kotak dengan lebar 5 dan tinggi 1, menggambarkan keseluruhan jarak kecuali jarak diagonal. Dalam ruang 3 dimensi kita dapat memodelkan semua jarak dengan tetrahedron. Tentu saja model ruang euclidean masih dapat dihitung dengan triangular, yang pada umumnya tidak cocok dengan jarak internet. Hasilnya menambahkan dimensi melebihi batas tertentu tidak membantu.

3. Mengurangi pemborosan pengukuran

Sejauh ini kita mengamsusikan bahwa setiap host akhir ahrus mengukur jaraknya dengan semua host landmark dalam upaya menentukan koordinat. Bagaimanapun, hanya $D+1$ jarak host ke landmark yang benar-benar dibutuhkan untuk menghitung koordinat di ruang dimensi D . Untuk membuktikan, kita melakukan percobaan dengan 15 landmark dan 7 dimensi ruang euclidean, dimana secara acak memilih 8 dari 15 landmark untuk setiap host akhir untuk menentukan koordinat. Kita menemukan bahwa 90% eror relatif dari GNP bertambah dari 0.5 menjadi 0.65. Ketika kita memilih 8 landmark yang paling dekat dengan setiap host akhir untuk menentukan koordinat, prediksi akurasi tidak berubah. Saat pembelajaran lebih lanjut tentang teknik ini dibutuhkan, terlihat sangat mudah diterapkan untuk mengurangi pemborosan pengukuran tanpa mengorbankan akurasi.

4. Kenapa tidak koordinat geografis?

Kenapa GNP tidak menggunakan hubungan geografis dengan host akhir. Jika begitu maka langsung saja menggunakan koordinat geografis (bujur dan lintang) untuk memperkirakan jarak. Kita memperoleh koordinat geografis untuk probe dan target dari set data global yang berasal dari NetGeo [8]. Meskipun teknik yang lebih mutakhir sudah dikemukakan, tool NetGeo tersedia untuk publik dan kita menggunakannya sebagai perkiraan pertama. Kita menghitung koefisien korelasi linear antara jarak geografis dan jarak sesungguhnya, dan juga antara hasil GNP dan jarak sesungguhnya, tidak termasuk jarak sesungguhnya yang lebih dari 2500ms, overall korelasi antara jarak geografis dan jarak terukur adalah 0.683, sedangkan korelasi antara jarak GNP dan jarak terukur adalah 0.915. mengetahui tool NetGeo tidak 100% akurat, dengan hati-hati perbedaan performa antara jarak GNP dengan jarak geografis membawa kita untuk percaya bahwa GNP menemukan relasi dalam spesifikasi jaringan melebihi relasi geografis.

V. KESIMPULAN

Setelah memodelkan internet sebagai ruang geometri dan mengkarakterisasi posisi setiap host di internet dengan sebuah nilai di ruang ini. Yang kemudian diuji dengan menggunakan pendekatan triangulated heuristic, dan global network positioning dan membandingkan hasilnya dengan jarak sesungguhnya dan jarak dengan metode IDMaps, maka dapat disimpulkan bahwa :

1. Hasil dari prediksi jarak menggunakan pendekatan triangulated heuristic dan global network positioning lebih menunjukkan keakuratan dibanding dengan metode IDMaps
2. Pendekatan global network positioning menghasilkan prediksi jarak yang lebih tepat dengan jarak sesungguhnya dibanding pendekatan triangulated heuristic.
3. Menunjukkan bahwa internet atau dunia maya pun dapat diukur jaraknya secara digital dengan memodelkan internet sebagai ruang geometri.
4. Optimalisasi jaringan untuk sistem terdistribusi, terutama peer to peer sangat mungkin dilakukan.
5. Prediksi jarak dapat membantu kemampuan dalam menentukan lokasi sumber informasi, selain dari ketersediaan bandwidth dan latensi.

Proses optimalisasi jaringan peer to peer di internet perlu diterapkan dan dikembangkan lebih lanjut, karena merupakan solusi yang lebih murah dari memperbesar bandwidth. Penelitian ini masih dapat dikembangkan lebih lanjut untuk model client – server. Optimalisasi Pemilihan lokasi dan jumlah landmark masih belum diketahui, jadi masih terbuka penyempurnaannya

REFERENSI

- [1] Wikipedia, "Bandwidth (computing)," Wikipedia, 13 Februari 2014. [Online]. Available: [http://en.wikipedia.org/wiki/Bandwidth_\(computing\)](http://en.wikipedia.org/wiki/Bandwidth_(computing)). [Diakses 24 Juni 2014].
- [2] M. Syafrizal, Pengantar Jaringan Komputer, Yogyakarta: ANDI, 2005.
- [3] M. Sutiyadi, "Pengenalan Internet," Ilmukomputer.com, Indonesia, 2003.

- [4] D. Ardiyansah, "Teknologi Jaringan Komputer," Ilmukomputer.com, Indonesia, 2003.
- [5] "NetGeo - The Internet Geographic Database," CAIDA, 25 September 2013. [Online]. Available: <http://www.caida.org/tools/utilities/netgeo/>. [Diakses 24 Juni 2014].
- [6] H. S.M., "Routing information organization to support scalable interdomain," 1994.
- [7] N. Awaludin, Geographical Information Systems with ArcGIS 9.X, Indonesia: ANDI, 2010.
- [8] F. P., J. S., P. V., Z. L., G. D.F. dan J. Y., "An architectur for a global internet host distance estimation service," dalam IEEE INFOCOM '99, New York, 1999.
- [9] J. A. Nelder dan M. R., "A Simplex Method for Function Minimization," Computer Journal, vol. VII, pp. 308-313, 1965.
- [10] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt dan L. Zhang, "IDMaps: A Global Internet Host Distance," IEEE/ACM Transactions on Networking (TON), vol. 9, no. 5, pp. 525-540, 2001.
- [11] T. E. Ng dan H. Zhang, "Predicting Internet Network Distance with Coordinates-Based Approaches," vol. 1, pp. 170-179, 2002.
- [12] A. A. Pangera.M.Kom dan D. Ariyus, Sistem Operasi, Yogyakarta: ANDI, 2005.