Permodelan dan Analisis Performa Jaringan Antrian dari Sistem Rujukan Berjenjang BPJS di Surabaya

Khairurizal Alfathdyanto, Abdullah Alkaff, Nurlita Gamayanti
Jurusan Teknik Elektro, Fakultas Teknologi Industri, Institut Teknologi Sepuluh Nopember
Jl. Arief Rahman Hakim, Surabaya 60111
e-mail: khairurizal12@mhs.ee.its.ac.id, abealkaff@gmail.com, nurlitagamayanti@gmail.com .

***Abstrak* – Sistem rujukan berjenjang diterapkan oleh BPJS sebagai bagian dari prosedur penjaminan kesehatannya. Pasien yang ingin mendapatkan penjaminan kesehatan pada fasilitas kesehatan (faskes) tertentu harus mendapat rujukan dari faskes tingkat I pasien tersebut terdaftar. Studi lapangan menunjukkan bahwa terjadi penumpukan pasien pada faskes lanjutan tertentu. Hal ini disebabkan tidak ada aturan yang membatasi pemberian rujukan. Dalam tugas akhir ini akan dibuat model jaringan antrian dari sistem rujukan BPJS dan dianalisis performa antrian dalam jaringan tersebut. Sistem rujukan berjenjang dapat dimodelkan sebagai jaringan antrian yang tersusun dari sistem antrian masing-masing faskes. Melalui analisis performa jaringan antrian tersebut dapat ditinjau pengaruh pola rujukan terhadap penumpukan pasien pada faskes pilihan dan perlunya aturan dalam pemberian rujukan. Lalu, hasil rekomendasi akan dibandingkan dengan sistem yang ada saat ini melalui simulasi. Hasil yang didapatkan adalah penurunan deviasi workload pada routing dinamis apabila sistem tidak sibuk. Dan pengaruh prioritas pasien tidak ada apabila laju pelayanannya sama.**

***Kata kunci* – Sistem Rujukan Berjenjang, Deviasi Workload, Prioritas user**

I. PENDAHULUAN

Sistem rujukan berjenjang diterapkan oleh BPJS sebagai bagian dari prosedur penjaminan kesehatannya. Pasien yang ingin mendapatkan penjaminan kesehatan harus melalui proses pemeriksaan di faskes tingkat I. Pasien yang membutuhkan penanganan lebih lanjut dapat memperoleh rujukan menuju faskes lanjutan. Sistem yang diterapkan BPJS ini masih menimbulkan masalah seperti menumpuknya pasien di rumah sakit terkenal yang memiliki fasilitas lengkap. Tidak adanya aturan yang membatasi pemberian rujukan menyebabkan tingginya rasio rujukan. Sebagai contoh, Puskesmas 5 Ilir, Sumatera Barat, rasio rujukannya dapat mencapai 60%.[1] Hal ini juga dipengaruhi beberapa permintaan rujukan yang didasari keinginan pasien untuk dirujuk sehingga menambah rasio rujukan.

Kualitas dari pelayanan BPJS masih diragukan. Salah satu penyebabnya adalah alur rujukan berjenjang yang diterapkan BPJS masih belum dibatasi dengan ketat. Rujukan saat ini masih memberikan kebebasan pada pasien untuk memilih fasilitas kesehatan lanjutan. Hal ini mengakibatkan terjadinya penumpukan pasien di fasilitas kesehatan yang menjadi favorit karena fasilitas yang lebih lengkap maupun pelayanan yang lebih memuaskan. Analisis performa jaringan antrian yang dimodelkan dari sistem rujukan BPJS akan menunjukkan pengaruh rujukan yang tidak diatur terhadap penumpukan pasien di faskes lanjutan tertentu. Hasil dari analisis akan menunjukkan perlunya pemberian aturan yang dapat menyeimbangkan rasio rujukan sehingga tidak terjadi penumpukan pasien.

Penumpukan antrian pasien pada sistem rujukan berjenjang merupakan permasalahan jaringan antrian pada pelayanan publik.[2],[3] Permasalahan yang diambil pada tugas akhir ini antara lain penentuan *routing* pada sistem rujukan berjenjang yang memperhatikan preferensi dan prioritas pasien. Pada tugas akhir ini akan didapatkan model jaringan antrian yang merepresentasikan sistem rujukan berjenjang. Model jaringan tersebut dapat digunakan untuk menghitung performa antrian dari masing-masing unit pelayanan kesehatan. Kemudian model tersebnut akan diuji untuk berbagai

II. TEORI SISTEM ANTRIAN

A. Sistem Antrian M/M/n dan Performanya [4]

Sistem antrian secara umum terdiri atas tiga unsur yaitu, sumber user (populasi), antrian, dan server (dapat berisi satu atau lebih server identik). Pendeskripsian sistem antrian memerlukan informasi tentang tiga elemen dasar sistem antara lain, proses pembangkitan user (proses kedatAangan pengguna di sistem), disiplin antrian (urutan user dalam mendapatkan akses ke server) dan proses pelayanan server.

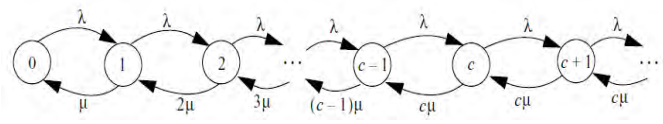
Performa sistem antrian secara umum dapat dilihat dari nilai rata-rata waktu tunggu dalam sistem, rata-rata waktu tunggu dalam antrian, rata-rata jumlah pelanggan dalam sistem, rata-rata jumlah pelanggan dalam antrian, proporsi jumlah pelanggan yang ditolak sistem, dan utilisasi server. Performa antrian dari suatu sistem antrian dapat dihitung dengan menggunakan nilai probabilatas dari masing-masing state sistem. Transisi state untuk model M/M/n diilustrasikan menggunakan diagram transisi state pada gambar 1.

Rata-rata banyak user yang mengantri dapat dihitung menggunakan persamaan (1),

$$E[N_q] = \frac{P_0 (n\rho)^n \rho}{n!\,(1-\rho)^2} \ldots (1)$$

Rata-rata waktu antri user dalam sistem dihitung menggunakan persamaan (2),[5]

$$E[W] = \frac{P_Q}{\lambda(1-\rho)} \dots (2)$$



Gambar 1. Diagram Transisi State untuk Model Antrian M/M/n

B. Hypercube Queueing System [6]

Hypercube queueing system (HQS) merupakan pengembangan dari spatial distributed queue (SDQ) untuk banyak server lebih dari 2. SDQ adalah model antrian dengan user maupun server terdistribusi secara spasial pada sistem. Area dari sistem dimodelkan menjadi sekumpulan *atom geografis* yang mewakili sumber permintaan pelayanan. Area respon utama adalah area dimana suatu unit ditugaskan ketika siap meskipun unit lain juga dalam keaadaan siap.

Ketersediaan server pada HQS dinyatakan sebagai state space sehingga untuk 3 server kondisi state membentuk suatu kubus dan untuk N-server, N>3, state akan membentuk hypercube. Server dalam model ini dapat diam maupun bergerak. Waktu tempuh antar atom perlu diperhatikan untuk tipe server yang bergerak. Namun, tipe server yang digunakan pada tugas akhir ini adalah tipe server diam.

Tanpa memperhatikan identitas server sistem antrian hypercube dapat dilihat sebagai sistem antrian M/M/n. Sistem dapat dilihat sebagai sistem antrian M/M/n apabila memenuhi asumsi-asumsi berikut antara lain,

1. Waktu antar kedatangan user memiliki distribusi eponensial dan hanya satu dari N server yang melayani user tersebut.
2. Waktu pelayanan semua server untuk tiap layanan memiliki distribusi eksponensial dengan rata-rata yang sama $(1/\mu)$.
3. Waktu pelayanan tidak dipengaruhi identitas server maupun user dan kondisi sebelumnya pada sistem.
4. Untuk kapasitas antrian tak berhingga, ketika semua server sedang sibuk user masuk dalam antrian yang sama dan kemudian akan dilayani dengan disiplin FCFS.

Preferensi user untuk memilih server yang melayaninya dapat mengakibatkan ketakseimbangan workload. Hal ini dikarenakan suatu server menjadi pilihan kebanyakan user dalam area tersebut. Ada beberapa cara untuk menyatakan ketakseimbangan workload, antara lain:

1. Deviasi absolut

$$\Delta W = \max_{i \neq j, i,j \in N} \left(|\rho_i - \rho_j|\right) \dots (3)$$

2. Varians

$$\sigma_W^2 = \sum_{i=1}^{N} \rho_i^2 - \left(\sum_{i=1}^{N} \rho_i\right)^2 \dots (4)$$

3. Standar deviasi

$$\sigma_W = \sqrt{\sum_{i=1}^{N} \rho_i^2 - \left(\sum_{i=1}^{N} \rho_i\right)^2} \dots (5)$$

4. Persentase deviasi maksimum terhadap mean

$$P = \left(\frac{\Delta W}{\bar{\rho}}\right) x100\% \dots (6)$$

C. Dynamic Routing [7]

Routing dinamis merupakan aturan yang menempatkan user pada server tertentu sesuai dengan kondisi server yang memenuhi kriteria tertentu. Routing dinamis dapat dibagi menjadi tiga komponen utama:

1. Information policy, yang memutuskan informasi state sistem harus dikumpulkan dan bagaimana informasi diperoleh.
2. Transfer policy, yang memutuskan apakah user yang datang layak untuk diproses sendiri atau dipindahkan.
3. Location policy, yang menentukan node yang layak untuk melayani user yang ditransfer.

Adaptive Virtual Delay (AVD) merupakan salah satu jenis routing dinamis. Kebijakan AVD bertujuan untuk menyeimbangkan delay virtual pada tiap server. Delay virtual tersebut terdiri atas delay virtual layanan dan delay virtual apabila terjadi transfer. Perkiraan delay virtual layanan pada server i dinotasikan sebagai F_i dan delay transfer virtual dinotasikan G. Nilai dari F_i dan G dapat dinyatakan sebagai berikut:

$$F_i = s_i(n_i + 1) \dots (7)$$

$$G_{ij} = \frac{t_{ij}}{1 - \rho'_{ij}} \dots (8)$$

Keterangan:

s_i = rata-rata waktu pelayanan server-i $(1/\mu_i)$

n_i = banyak antrian pada server-i

t_{ij} = rata-rata waktu tempuh user dari server-i ke server-j

ρ'_{ij} = tingkat kesibukan jalan antara server-i dan server-j

Delay virtual layanan merupakan total waktu user harus menetap pada server-i jika user itu memasuki server-i pada saat penghitungan. Sementara delay transfer virtual merupakan waktu tempuh antar server apabila user dipindahkan menuju server lain. Adapun deskripsi dari transfer policy dan location policy untuk AVD sebagai berikut,[8]

Transfer policy adalah pembandingan banyak antrian pada server terhadap batas banyak antrian. Setiap server menggunakan hanya informasi internal untuk menentukan user dilayani atau tidak. User akan dipindahkan menuju server lain apabila banyak antrian pada server tersebut melebihi suatu batasan (T_i) dan akan tetap dilayani server tersebut apabila dibawah T_i.

Penentuan nilai T_i pada metode AVD pertama-tama ditentukan T_j sebagai batasan untuk server tercepat-j. Dengan mengganti variabel n_j dan n_i dengan T_j dan T_i secara berturut-turut pada persamaan (7), maka akan diperoleh nilai batasan untuk semua server.

$$n_i = \left(\frac{s_i}{s_j}\right)(n_i + 1) - 1 \dots (9)$$

Location policy merupakan prosedur yang digunakan untuk menentukan server yang melayani user apabila dipindah tempatkan. Prosedur location policy dalam AVD antara lain

1. Cari server dengan beban paling ringan.

User yang datang saat antrian di server-i sedang melebihi batasan. Pada saat tersebut nilai beban di server-i dan server kandidat tujuan j ($j \in N, j \neq 1$) dengan aturan berikut

- Jika $F_i > F_j + G_{ij}$, maka server i memiliki lebih sibuk daripada server-j.
- Sebaliknya jika $F_i < F_j + G$, maka server i memiliki tidak lebih sibuk daripada server-j.

Kemudian misalkan $\delta_{ij} = F_i - (F_j + G)$ dan $\delta_i = \max(\delta_{ij})$ Jika δ_i lebih besar dari nol, server-j dengan $\delta_{ij} = \delta_i$ adalah node dengan beban terendah.

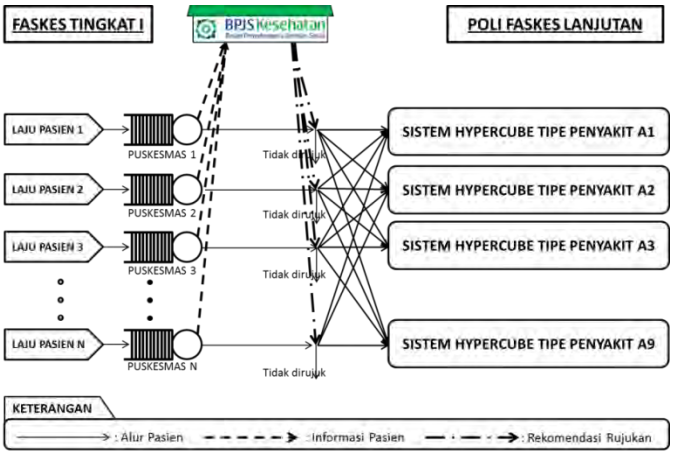
2. Mengalokasikan user

Definisikan w sebagai faktor pembobot alokasi dan Δ sebagai nilai bias. Jika $w^c\delta_i > \Delta$ maka user akan dipindah menuju server-j sebanyak c kali. Faktor pembobot w digunakan untuk membatasi pemindahan user. Nilai bias Δ digunakan untuk melindungi sistem dari potensi ketidakstabilan.

III. METODE PERMODELAN

A. Skema Permodelan Jaringan

Jaringan antrian adalah kumpulan sistem antrian yang saling terhubung. Perhatikan bahwa tiap faskes dalam sistem pelayanan kesehatan dapat ditinjau sebagai suatu sistem antrian yang terhubung melalui sistem rujukan berjenjang. Peserta JKN yang dirujuk dari faskes I akan melanjutkan antrian pada faskes tingkat lanjutan. Berdasarkan fakta tersebut sistem rujukan berjenjang dapat dimodelkan sebagai suatu jaringan antrian. Pembahasan model jaringan antrian dari sistem rujukan berjenjang dibagi menjadi 3 sub-bagian antara lain sistem antrian M/M/n/-/P untuk fasilitas kesehatan tingkat I, proporsi rujukan yang dimodelkan dengan metode BCMP, dan perujukan pasien dengan menggunakan model antrian hypercube. Hal tersebut digambarkan dalam skema di gambar 2.



Gambar 2. Skema Permodelan Sistem Rujukan Berjenjang

B. Model Antrian Hypercube

Notasikan H_i $(i = 1,2,3,\ldots,9)$ sebagai model hypercube untuk penyakit A_i. Misalkan λ_i adalah laju kedatangan keseluruhan User pada tiap model adalah pasien rujukan dari faskes I yang telah dikelompokkan berdasarkan jenis penyakitnya. Untuk penyakit ke-i dan atom geografis k, laju kedatangan pasien dinotasikan sebagai λ_{ik} sehingga nilai λ_i dapat dituliskan sebagai berikut,

$$\lambda_i = \sum_{k=1}^{18} \lambda_{ik} \ldots (10)$$

Server pada tiap model adalah dokter yang menangani jenis penyakit tersebut pada tiap poli di rumah sakit. Notasikan n_{il} adalah banyak dokter yang melayani penyakit i di rumah sakit l. Maka, nilai n_i yang mewakili banyak server pada H_i adalah,

$$n_i = \sum_{l=1}^{3} n_{il} \ldots (11)$$

Rumah sakit yang tidak memiliki dokter yang dapat melayani jenis penyakit-i tidak dimasukkan dalam model penyakit ke-i.

Pasien ditempatkan pada suatu server berdasarkan tingkat preferensi yang dinyatakan dalam matriks preferensi. Notasikan matriks preferensi sebagai P_i. Matriks preferensi pada model H_i merupakan matriks berukuran $18xn_i$. Pada model hypercube matriks preferensi berisi nilai preferensi server terhadap user. Namun dengan membalikkan alur penempatan pada model sistem rujukan ini, matriks preferensi berisi nilai urutan tingkat preferensi pasien terhadap server di rumah sakit. Nilai tersebut akan didapatkan dari proporsi rujukan yang berasal dari data rujukan di BPJS pada tahun 2014. Pembagian wilayah Surabaya Timur digambarkan pada gambar 3.

Gambar 3. Skema Pembagian Atom Geografis dan Posisi Server

C. Dynamic Routing untuk Sistem Rujukan

Tiga komponen dynamic routing untuk sistem rujukan didefinisikan sebagai berikut

1. Information policy; Informasi yang diambil adalah panjang antrian pada server tiap waktu. Pengambilan informasi panjang antrian diasumsikan mengalami waktu delay sebesar. Estimasi laju kedatangan tiap server diambil dari hasil perhitungan hypercube.
2. Transfer policy; Pasien dirujuk menuju server preferensi utama, selanjutnya apabila preferensi utama memiliki antrian yang padat maka pasien dirujuk ke server lain
3. Position policy; Pasien dirujuk menuju server dengan nilai fungsi preferensi paling besar

Fungsi preferensi yang diusulkan untuk routing dinamis didapatkan dari ekspektasi waktu tunggu untuk server tujuan. Nilai ini didapat pada persamaan (2). Kemudian diberikan sedikit modifikasi yaitu penambahan ekspektasi waktu tunggu sesuai dengan ekspektasi bertambahnya antrian saat menempuh perjalanan menuju server. Kemudian diasumsikan bahwa tiap pasien rela menunggu T jam lebih lama apabila berada pada faskes dengan tingkat preferensi yang lebih tinggi menurut user tersebut. Fungsi preferensi dinyatakan dalam persamaan berikut,

$$F_{ik} = ((n_k + \lambda_k * T_{ik})/\mu_k + P_{ik} * T)^{-1} \dots (12)$$

Dimana:
i = indeks atom geografis
k = indeks server
n_k = banyak antrian pada server k
P_{ik} = urutan preferensi server k menurut user dari atom i
T_{ik} = waktu tempuh dari atom i ke server k

D. Prioritisasi Pasien

Pasien dikelompokkan menjadi 2 kelompok prioritas sesuai tingkat kegawatan penyakit yaitu kronis dan non-kronis. Pasien kronis memiliki waktu pelayanan yang lebih lama daripada pasien non-kronis. Pasien kronis akan mendahului

Notasikan p_0 sebagai proporsi pasien dengan penyakit kronis. $\bar{\mu}$ merupakan laju pelayanan rata-rata kesuluruhan pasien. Proses pelayanan pasien kronis juga merupakan proses Poisson dengan laju $\mu_0 = k.\bar{\mu}$. Nilai laju pelayanan harus dijaga konstan sehingga μ_1 yang merupakan laju pelayanan pasien non-kronis harus memenuhi persamaan (13)

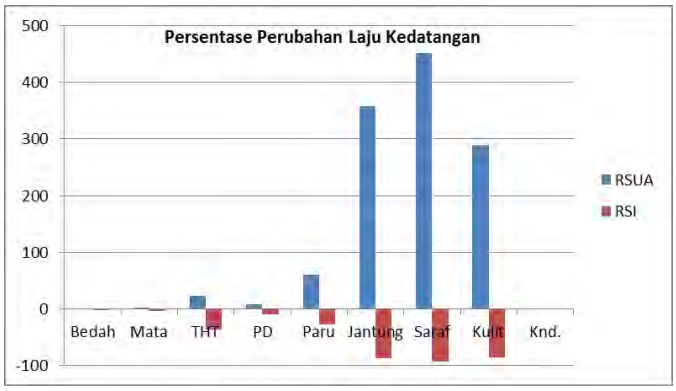
$$\mu_1 = \frac{(1 - p_0)k\bar{\mu}}{k - p_0} \dots (13)$$

IV. HASIL PENGUJIAN DAN ANALISIS

Model sistem rujukan disimulasikan menggunakan toolbox SimEvent pada Simulink. Parameter yang dipergunakan berasal dari pengolahan data BPJS pada tahun 2015. Laju keberangkatan rujukan menuju 3 rumah sakit dalam model dari Surabaya pada tahun 2015 ditunjukkan pada tabel 1. Parameter lain yang dipergunakan adalah laju pelayanan yang didapatkan dari survei dan waktu tempuh menuju rumah sakit yang diolah dengan bantuan Google Maps.

A. Analisis Sensitivitas terhadap Preferensi

Pengujian dilakukan dengan mengubah preferensi RSUA menjadi 1 tingkat diatas RSI menurut pasien dari area Gading, Rangkah, Pacar Keling, Kalijudan, Mojo, Mulyorejo, Keputih dan Klampis Ngasem. Pengubahan preferensi ini berdasarkan jarak tempuh menuju rumah sakit. Perubahan proporsi routing terjadi pada sistem di masing-masing poli penyakit dan ditunjukkan pada Gambar 4.

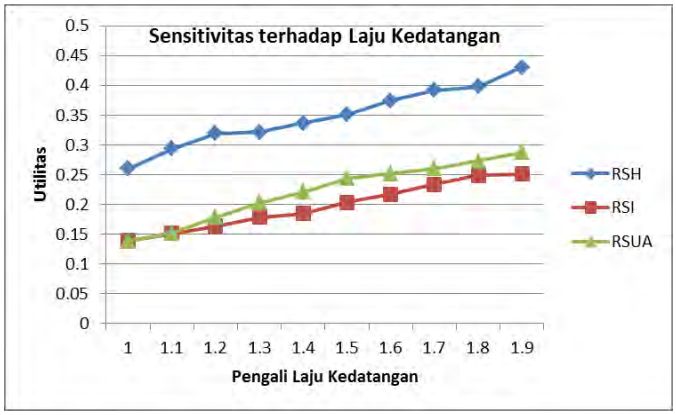


Gambar 4. Perubahan Laju Kedatangan

Laju kedatangan di RSUA mengalami peningkatan dan sebaliknya untuk RSI. Pengaruh perubahan preferensi ini berhubungan dengan utilisasi pada sistem. Perubahan laju kedatangan RSUA lebih kecil pada sistem yang sibuk atau utilisasinya besar dan sebaliknya. Hal ini dapat diinterpretasikan bahwa semakin ramai sistem pasien tidak dapat dengan leluasa memilih rumah sakit tujuan.

B. Sensitivitas terhadap Laju Kedatangan Global

Pengujian dilakukan dengan meningkatkan semua laju kedatangan secara bertahap yaitu 10% pada tiap tingkatan. rumah sakit. Perubahan utilisasi terjadi pada sistem di masing-masing poli pada tiap rumah sakit. Gambar 5 menunjukkan perbandingan utilisasi poli penyakit Paru dari 3 rumah sakit pada model.

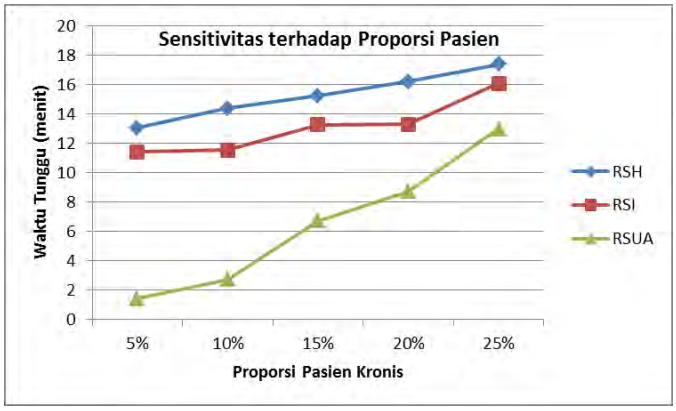


Gambar 5. Sensitivitas Routing Dinamis terhadap Laju Kedatangan

Utilisasi poli semakin meningkat dengan bertambahnya laju kedatangan. Kendati demikian, rasio utilisasinya tetap pada 45:25:30. RSUA mengalami peningkatan lebih tinggi daripada RSI karena memiliki 3 dokter penyakit paru sementarra RSI hanya memiliki 1 meskipun RSI memiliki tingkat preferensi lebih tinggi. Hal ini karena RSUA lebih mampu daripada RSI

C. Sensitivitas terhadap Prioritas Pasien

Pengujian dilakukan dengan meningkatkan proporsi pasien kronis secara bertahap 5% di setiap tingkatan. Perubahan waktu tunggu diamati pada penyakit bedah yang memiliki pasien lebih banyak dan tingkat kesibukan tinggi. Gambar 6 menunjukkan perbandingan waktu tunggu poli penyakit Bedah dari 3 rumah sakit pada model.



Gambar 6. Sensitivitas Routing Dinamis terhadap Proporsi Pasien Kronis

Peningkatan pasen kronis meningkatkan waktu tunggu pada tiap poli. Gambar 6 menunjukkan peningkatan yang besar pada RSUA. Hal ini menunjukkan efektifiktas routing dinamis yang mengarahkan pasien menuju rumah sakit dengan waktu tunggu yang lebih rendah.

V. KESIMPULAN

Setelah melakukan analisis hasil simulasi model untuk routing statis, dinamis dan pemberian prioritas pelayanan pasien, penulis dapat menyimpulkan hal-hal berikut:

1. Pengubahan preferensi akan mengubah laju kedatangan menuju rumah sakit sesuai urutan preferensinya.
2. Penerapan routing dinamis dapat memperbaiki kondisi tersebut sehingga senua antrian stabil.
3. Penerapan routing dinamis memiliki pengaruh yang besar terhadap waktu tunggu dan workload pada sistem dengan workload awal yang tinggi.
4. Peningkatan laju kedatangan meningkatkan utilisasi tiap rumah sakit, pengaruh routing dinamis mengakibatkan lebih pasien banyak dirujuk menuju rumah sakit dengan dokter yang lebih banyak.
5. Terjadi perubahan waktu tunggu pada model saat pemberian prioritas pasien dibandingkan saat hanya menggunakan routing dinamis.
6. Penerapan prioritas pasien akan memperbaiki routing dinamis pada kondisi workload sistem yang tinggi.
7. Routing dinamis mendistribusi pasien sesuai waktu tunggu sehingga ketika jumlah pasien kronis bertambah rumah sakit yang lebih sepi terisi.

DAFTAR PUSTAKA

[1] Dede S. 2015, „Analisis Pelaksanaan Sistem Rujukan Rawat Jalan Tingkat Pertama (RJTP) pada Peserta BPJS Kesehatan di Puskesmas 5 Ilir dan Puskesmas Merdeka", Repositori Universitas Sriwijaya.

[2] Mengyu G., dkk. 2011, „Effectiveness of Referral Incentive Policy: Exploring Using Queueing Network Model with Blocking", dalam *International Conference on Service Systems and Service Management (ICSSSM), 2011*, hal. 1-6.

[3] Cayirli, T. dkk. 2006, „Designing Appointment Scheduling Systems for Ambulatory Care Services", *Journal of Health Care Management Sciences*, no. 9, hal. 47-58.

[4] Trivedi, K.S. 2001, *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*, 2nd Ed, Wiley, New Jersey.

[5] Oliver, C.I. 2011, *Fundamentals of Stochastic Network,* Wiley, New Jersey.

[6] Larson, R.C. 1981, *Urban Operation Research,* Prentice-Hall, New Jersey.

[7] Zhang Y., dkk., "A Performance Comparison of Adaptive and Static Load Balancing in Heterogeneous Distributed Systems," *Proc. IEEE 28th Ann. Simulation Symp.*, pp. 332-340, Phoenix, Ariz., April 1995.

[8] Watts, J. dan Taylor, S. "A Practical Approach to Dynamic Load Balancing," *IEEE Trans. Parallel and Distributed Systems*, vol. 9, no. 3, pp. 235-248, March 1998.

[9] Larson, R.C. 1974, *Urban Emergency Service System: An Iterative Procedure for Approximating Performance Characteristics,* The Rand Institute, New York.