

# Analisis dan Pemodelan Trafik Internet Spasial

Sis Soesetijo<sup>1\*</sup>, Febrianto Budimulyono<sup>2\*\*</sup>, Lukas Hadi Purnama<sup>3\*\*</sup>,  
Welly Wellandow Santoso<sup>4\*\*</sup>, Hendrik Setiawan<sup>5\*\*</sup>

<sup>\*)</sup> Jurusan Teknik Elektro Universitas Surabaya, Surabaya 60293, email: <sup>1)</sup>[ssoesetijo@ubaya.ac.id](mailto:ssoesetijo@ubaya.ac.id)

<sup>\*\*)</sup> Jurusan Teknik Elektro Universitas Surabaya, Surabaya 60293,  
email:<sup>2)</sup> [febrianto\\_b4130n@hotmail.com](mailto:febrianto_b4130n@hotmail.com), <sup>3)</sup>[squall\\_fire2000@yahoo.com](mailto:squall_fire2000@yahoo.com),

<sup>4)</sup> [w\\_ly\\_inzsight@hotmail.com](mailto:w_ly_inzsight@hotmail.com), <sup>5)</sup> [hendriksetiawan@gmail.com](mailto:hendriksetiawan@gmail.com)

**Abstrak** – Pengukuran trafik internet spasial dilakukan secara agregat selama satu tahun dengan mengambil 4 lokasi pengukuran trafik di kampus Universitas Surabaya yaitu Fakultas Bisnis dan Ekonomika, Fakultas Teknik, Perpustakaan dan Kampus Ubaya Ngagel. Pemodelan trafik menggunakan Vector Auto Regressive (VAR) dengan validasi model menggunakan qqplot dan uji distribusi normal pada residu model. Hasilnya diperoleh bahwa model VAR(1) merupakan model yang sesuai untuk data trafik internet spasial tersebut. Untuk mengetahui hubungan kausalitas antar trafik dilakukan analisis kausalitas menggunakan metoda Granger Causality. Hasil analisisnya menunjukkan bahwa trafik internet di Perpustakaan dipengaruhi oleh ketiga trafik yang lainnya.

**Kata Kunci:** Trafik Internet Spasial, VAR, qqplot, Granger Causality, Trafik Agregat

## 1. PENDAHULUAN

Pertumbuhan trafik internet di Indonesia selama 5 tahun terakhir terbilang sangat tinggi. Bahkan para operator telekomunikasi, saat ini lebih cenderung menjual layanan data daripada layanan suara. Untuk mengetahui karakteristik trafik internet secara lebih mendalam harus dilakukan pemodelan terhadap data trafik internet sehingga model tersebut dapat digunakan untuk evaluasi dan prediksi trafik internet. Salah satu pemodelan yang paling banyak digunakan adalah model ARIMA [1]-[3]. Pada makalah [1] pemodelan ARIMA untuk trafik internet dengan menggunakan data trafik internet secara agregat selama periode pengukuran 14 hari. Sedangkan pada makalah [2] periode pengukuran trafik internet agregat hanya 15 menit dengan membedakan antara trafik internet yang terserang hacker (virus) dengan yang tidak. Pada makalah [3], pemodelan trafik internetnya menggunakan model ARIMA yang lebih sederhana dan cepat yaitu model Adjusted ARIMA. Ketiga model di atas merupakan model ARIMA hanya untuk data trafik internet univariate saja. Sedangkan penelitian pemodelan trafik internet sebagai fungsi spasial dan temporal telah dilakukan dan dipresentasikan pada makalah

[4] dan [5]. Makalah [4] menggunakan data trafik Round Trip Time (RTT) pada beberapa router dan dianalisis menggunakan spasial korelasi sedangkan pada makalah [5] menggunakan data trafik TCP dan UDP.

Pada penelitian ini dibahas pemodelan deret waktu multivariate trafik internet agregat sebagai fungsi spasial selama periode pengukuran satu tahun menggunakan model Vector Auto Regressive (VAR). Pengukuran trafik internet dilakukan di kampus Universitas Surabaya. Untuk mewakili fungsi spasial, pengukuran trafik internetnya dilakukan pada 4 lokasi yaitu trafik pada router di fakultas teknik (E), fakultas bisnis dan ekonomika (M), perpustakaan (S) dan kampus Ubaya Ngagel (N).

Pemodelan VAR(p) pada makalah ini menggunakan metoda Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC) untuk estimasi nilai lag p. Pada penelitian ini digunakan nilai p = 1 dari hasil estimasi dengan BIC, karena pada uji *asymptotic* diperoleh nilai p-value yang paling kecil. Untuk memperoleh hubungan kausalitas antar trafik di masing-masing lokasi, dilakukan uji kausalitas menggunakan metoda Granger Causality. Uji ini untuk mencari hubungan sebab-akibat (*causalities*) antara trafik pada satu lokasi dengan lokasi yang lain. Apakah masing-masing variabel trafik di masa lampau berpengaruh signifikan atau tidak terhadap variable trafik di lokasi yang lain, atau hanya berpengaruh pada dirinya sendiri saja.

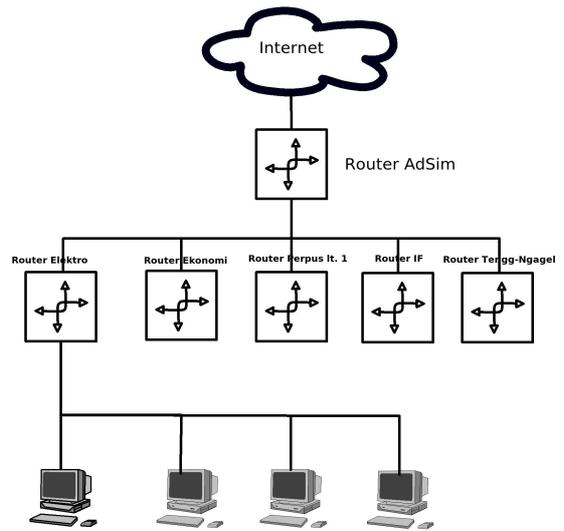
Metoda Granger Causality sudah berhasil diaplikasikan pada beberapa bidang antara lain pada bidang biomedik seperti pada makalah [6], bidang ekonomi pada makalah [7] dan bidang energi pada makalah [8].

## 2. METODOLOGI

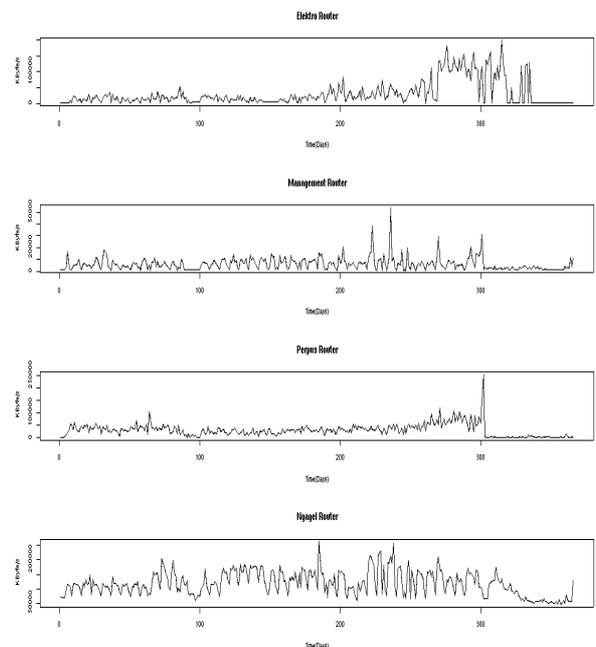
### 2.1. Sistem Pengukuran

Metoda yang digunakan dalam penelitian ini dibagi atas beberapa tahap seperti yang terlihat pada diagram alir metodologi penelitian pada gambar 1 di bawah ini. Langkah awal pada metoda penelitian adalah pengukuran trafik internet secara agregat pada 4 lokasi router di kampus Universitas Surabaya. Keempat lokasi router tersebut adalah

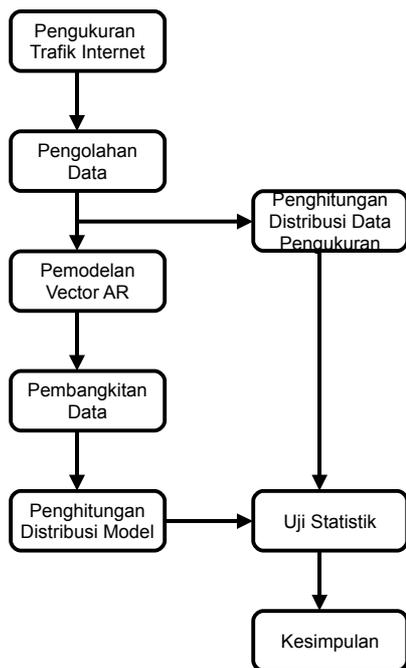
router fakultas teknik (E), router fakultas bisnis dan ekonomika (M), router perpustakaan (S) dan router kampus Ubaya Ngagel (N). Pemilihan keempat router diasumsikan dapat mewakili fungsi spasial dari keseluruhan trafik internet di kampus Ubaya. Topologi jaringan internet di kampus Ubaya ditunjukkan pada gambar 2. Untuk mendapatkan data trafik internet agregat harus dilakukan pengambilan data trafik dari router dengan menggunakan Simple Network Management Protocol (SNMP). Pengambilan data trafik internet menggunakan SNMP yang tersimpan dalam bentuk format logfile. Data trafik diambil dari bulan September 2009 sampai bulan September 2010 dengan menggunakan rata-rata harian (daily average). Trafik internet agregat yang diukur merupakan trafik *downstream* saja. Hasil pengukuran trafik internet dalam bentuk runtun waktu ditunjukkan pada gambar 3 berikut. Setelah semua proses pengambilan dan pengolahan data selesai, kemudian dilakukan pemodelan VAR menggunakan perangkat lunak statistik open source R. Sebelumnya data diuji terlebih dulu apakah sudah stasioner atau belum. Langkah ini perlu dilakukan untuk menghindari model yang bersifat heteroskedatik. Untuk menguji apakah model yang dihasilkan sudah sesuai maka dilakukan validasi model dengan menggunakan metoda qqplot yaitu perbandingan distribusi model dengan data pengukuran data trafik dan uji distribusi normal pada residu model. Langkah terakhir adalah melakukan analisis terhadap model VAR yang telah diperoleh menggunakan metoda Granger Causality untuk mengetahui hubungan kausalitas antar trafik internet terhadap keempat lokasi router.



Gambar 2. Topologi Jaringan Internet



Gambar 3. Hasil Pengukuran Trafik Internet



Gambar 1. Metodologi Penelitian

## 2.2. Model VAR

Bentuk dasar dari VAR terdiri dari  $K$  variabel  $y_t = (y_{1t}, \dots, y_{kt}, \dots, y_{Kt})$  untuk  $k = 1, \dots, K$ . Proses  $VAR(p)$  didefinisikan sebagai :

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + CD_t + u_t \quad (1)$$

di mana  $A_i$  adalah matrik koefisien ( $K \times K$ ) untuk  $i = 1, \dots, p$  dan  $u_t$  merupakan dimensi  $K$  dari proses white noise yang mempunyai matrik kovarian  $E(u_t u_t') = \Sigma_u$ . Matrik  $C$  adalah matrik koefisien yang berdimensi

$(K \times M)$  dan  $D_t$  merupakan matrik  $(M \times 1)$ . Persamaan (2.1) dapat ditulis dalam bentuk polinomial lag  $A(L) = (I_K - A_1 - \dots - A_p)$  sebagai berikut :

$$A(L) y_t = CD_t + u_t \quad (2)$$

Salahsatu karakteristik dari proses  $VAR(p)$  adalah stabilitasnya. Artinya bahwa prosesnya menghasilkan deret waktu yang stasioner dengan rata-rata yang tidak berubah pada fungsi waktu. Untuk mendapatkan parameter orde lag  $p$  dari  $VAR(p)$  dan koefisien  $A$  dari model, berikut langkah-langkah untuk identifikasi dan estimasinya, prosedur ini secara umum sama dengan model ARMA :

a. Uji Stasioner yaitu uji untuk melihat apakah data yang akan diamati stasioner atau tidak. Apabila belum stasioner maka perlu dilakukan penurunan pertama (*first differential*) agar diperoleh stasioner pada orde pertama. Uji stasioner yang umum digunakan adalah Augmented Dickey-Fuller (ADF) Test, dengan persamaan sebagai berikut :

$$\Delta y_t = \phi y_{t-1} + \sum_{j=1}^{p-1} A_j^* \Delta y_{t-j} + u_t \quad (3)$$

dengan  $\phi = -A(1)$  dan  $A_j^* = -(A_{j+1} + \dots + A_p)$ .

Dengan model persamaan ini, hipotesis yang digunakan adalah  $H_0: \phi=0$  versus  $H_1: \phi < 0$ . Jika hipotesis nol,  $H_0$  tertolak maka dipastikan data yang diuji adalah stasioner.

b. Proses Estimasi yaitu proses untuk menentukan koefisien model yang diamati. Penentuan koefisien  $VAR(p)$  yang paling efektif adalah dengan menggunakan *least-squares* yang diterapkan terpisah pada masing-masing persamaan.

c. Akaike Information Criterion (AIC) dari masing-masing regresi untuk menjamin bahwa residual yang dihasilkan bersifat White Noise. Uji ini untuk menentukan lag mana yang paling relevan dipakai dalam model.

### 2.3. Granger Causality

Tujuan dari uji kausalitas menggunakan Granger Causality adalah mendeteksi ada tidaknya hubungan sebab-akibat (causalities) antara variabelnya. Variabel  $x$  dikatakan "*granger-causes*" variabel  $y$ , apabila variabel  $x$  membantu memprediksi variable  $y$ . Untuk melakukan kedua uji tersebut, vector dari variabel  $y_t$  dibagi 2 subvektor  $y_{1t}$  dan  $y_{2t}$  dengan dimensi  $(K_1 \times 1)$  dan  $(K_2 \times 1)$  dengan  $K = K_1 + K_2$ . Proses  $VAR(p)$  dapat dituliskan kembali :

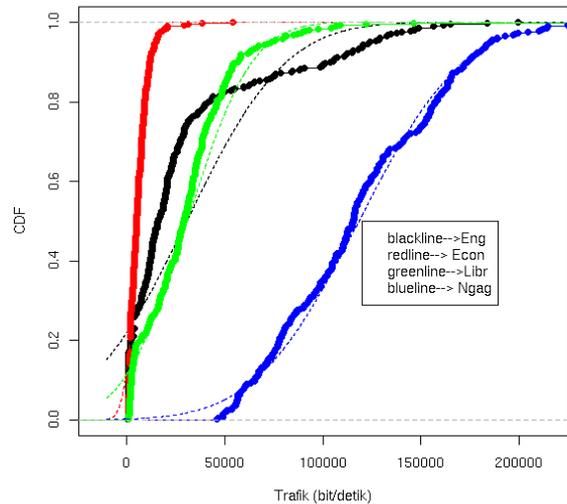
$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \sum_{i=1}^p \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} \\ \alpha_{21,i} & \alpha_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + CD_t + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \quad (4)$$

Subvektor  $y_{1t}$  tidak "*granger-cause*"  $y_{2t}$  apabila  $\alpha_{21,t} = 0$  untuk  $i = 1, 2, \dots, p$  dan sebaliknya Subvektor  $y_{1t}$  "*granger-cause*"  $y_{2t}$  apabila  $\alpha_{21,t} \neq 0$  untuk  $i = 1, 2,$

$\dots, p$ . Dengan demikian analisis Kausalitas akan menguji apakah sebuah variabel dapat membantu memprediksi variabel yang lain. Uji Granger Causality paling mudah diperoleh dari model  $VAR(p)$  [9].

## 3. HASIL DAN PEMBAHASAN

Pada makalah ini hasil yang dibahas dan dianalisis adalah trafik internet agregat *downstream* pada 4 lokasi router di kampus Universitas Surabaya pada periode September 2009 sampai dengan September 2010. Keempat lokasi router tersebut adalah fakultas Teknik (E), fakultas Bisnis dan Ekonomika (M), Perpustakaan (S) dan kampus Ubaya Ngagel (N). Analisis menggunakan software statistik open source R v2.12.2 pada sistem operasi Mandriva Linux 2008. Gambar 4 menunjukkan distribusi trafik internet agregat pada masing-masing lokasi trafik dan nampak bahwa trafik di kampus Ubaya Ngagel (N) jauh lebih besar dari ketiga yang lainnya.



Gambar 4. Distribusi Kumulatif Trafik Internet

### 3.1. Uji Stasioner Data Trafik

Untuk memperoleh model  $VAR(p)$  yang memiliki stabilitas yang baik, yaitu dalam proses pemodelannya menghasilkan deret waktu yang stasioner dengan rata-rata yang tidak berubah pada fungsi waktu. Oleh karena itu diperlukan uji stasioner terhadap data trafik pada masing-masing lokasi router. Uji stasioner menggunakan uji Augmented Dickey-Fuller (ADF).

Tabel 1 menunjukkan bahwa data trafik dengan uji ADF merupakan data stasioner dengan tingkat signifikansi 95% karena absolut *test-value* lebih besar dari absolut critical value 5% (ditandai \* pada tabel 1).

Dari uji stasioner ini menunjukkan bahwa data trafik sudah stasioner sehingga dapat dimodelkan  $VAR(p)$  dengan mencari terlebih dahulu nilai lag  $p$  dengan menggunakan estimasi AIC dan BIC.

Tabel 1. Uji Stasioner Data Trafik

Lokasi Trafik	Test-value	Critical Value	
		5%	10%
Teknik	-4.3078*	-1.95*	-1.62
Ekonomi	-5.703*	-1.95*	-1.62
Perpus	-3.800*	-1.95*	-1.62
Ngagel	-2.1476*	-1.95*	-1.62

### 3.2. Model VAR(1)

Hasil estimasi lag  $p$  menggunakan metoda AIC dan BIC masing-masing memperoleh nilai  $p = 8$  dan  $p = 1$ . Nilai lag  $p$  tersebut akan menghasilkan model VAR dengan nilai kesalahan (RMSE) yang kecil. Untuk menentukan nilai lag  $p$  yang digunakan pada makalah ini maka digunakan uji *asymptotic*. Hasil uji dengan nilai  $p$ -value yang terkecil itulah yang akan digunakan untuk pemilihan lag  $p$  seperti ditunjukkan pada tabel 2 berikut.

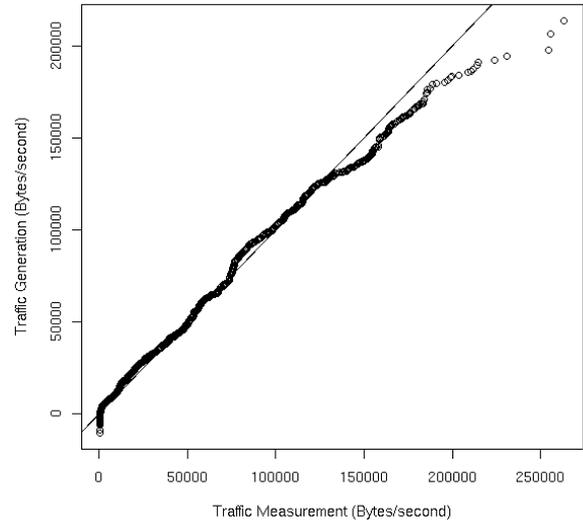
Tabel 2. Uji Asymptotic

Lag $p$	Chi-squared	p-value
$p = 1$	607.56	$2.2 \cdot 10^{-16}$
$p = 8$	210.19	$6.37 \cdot 10^{-6}$

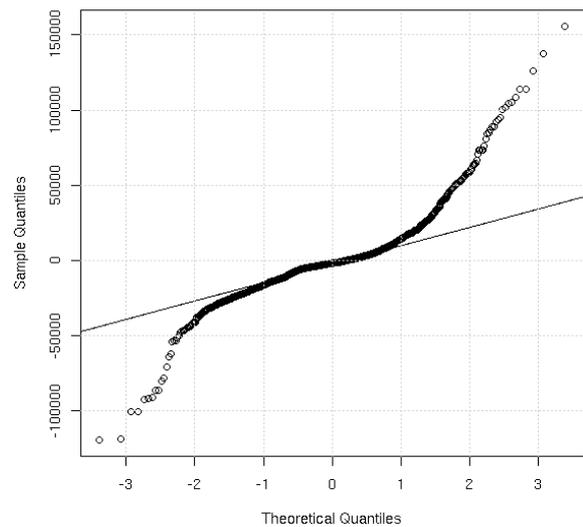
Hasil uji di tabel 2 tersebut menunjukkan bahwa model VAR(1) merupakan model terbaik yang digunakan pada pemodelan data trafik bulan September 2009 – September 2010.

Untuk mengetahui apakah model VAR(1) merupakan model yang sesuai dan valid untuk pemodelan data trafik tersebut, maka harus dilakukan validasi model menggunakan metoda qqplot. Metoda ini dipilih karena menggunakan perbandingan distribusi data pembangkitan model VAR(1) dengan data trafik hasil pengukuran. Gambar 5 menunjukkan hasil validasi menggunakan qqplot. Nampak bahwa perbandingan distribusi antara data model dengan data pengukuran mengikuti garis diagonal sampai data trafik 150 kbit/detik dan mulai posisi trafik 150 kbit/detik tersebut nampak perbandingan distribusi tersebut semakin menjauh. Keadaan tersebut menunjukkan bahwa pada trafik di atas 150 kbit/detik tersebut sudah mulai terdapat banyak data trafik yang outlier. Pada pemodelan VAR ini tidak dilakukan deteksi outlier. Namun demikian dapat disimpulkan bahwa hasil validasi model VAR(1) merupakan model yang sesuai dan valid untuk pemodelan data trafik spasial tersebut

karena distribusi residu model VAR(1) menunjukkan bahwa residu model memiliki distribusi normal dan mempunyai nilai rata-rata sama dengan nol seperti ditunjukkan pada gambar 6 berikut ini.



Gambar 5. Grafik qqplot untuk validasi model



Gambar 6. Uji Distribusi Normal model VAR(1)

### 3.3 Analisis Kausalitas

Analisis ini untuk mencari hubungan timbal balik (*interrelationship*) antara variabel trafik pada satu lokasi dengan lokasi trafik yang lain. Apakah masing-masing variabel trafik di masa lampau berpengaruh signifikan atau tidak terhadap variable trafik di lokasi yang lain, atau hanya berpengaruh pada dirinya sendiri saja. Uji ini akan diterapkan dengan menggunakan hasil orde lag  $p = 1$  pada model VAR(1).

Analisis ini bertujuan untuk mempelajari dan menunjukkan bagaimana formulasi matematik dan statistik dapat digunakan untuk mengungkap hubungan sebab-akibat yang terjadi pada trafik masing-masing lokasi. Analisis kausalitas paling mudah dan sederhana diaplikasikan pada model VAR(1). Kemudian dari model VAR(1), dicari

kemungkinan hubungan kausalitas yang terjadi dengan menggunakan Uji Statistik F (F-Test). Dari nilai nilai F-test tersebut dapat diperoleh nilai *p-value*. Oleh karena dalam uji sebelumnya dipilih nilai confidence interval 95%, maka nilai *p-value* harus lebih kecil dari 5% dengan hipotesa awal  $H_0$  : x tidak *granger-cause* terhadap y. Kekuatan kausalitas dapat diukur dari besarnya nilai F-test [8].

Hasil uji kausalitas ditunjukkan pada tabel 3 berikut ini. Nampak bahwa yang memenuhi terjadinya *granger-cause* hanya  $E \rightarrow S$ ,  $M \rightarrow S$ ,  $N \rightarrow S$  dan  $N \rightarrow M$  (pada tabel 3 diberi tanda \*).  $E \rightarrow S$  berarti bahwa trafik internet pada fakultas teknik (E) mempengaruhi trafik pada Perpustakaan (S). Yang menarik adalah bahwa trafik internet pada Perpustakaan (S) dipengaruhi oleh ketiga trafik lainnya yaitu Fakultas Teknik (E), Fakultas Bisnis dan Ekonomika (M) dan kampus Ubaya Ngagel (N). Trafik dari Fakultas Teknik [E] mempunyai pengaruh yang paling besar terhadap trafik Perpustakaan (S) dibandingkan dengan kedua trafik lainnya ( M dan N). Selain ketiga kausalitas di atas juga terjadi kausalitas searah dari trafik kampus Ubaya Ngagel (N) ke trafik Fakultas Bisnis dan Ekonomika (M). Kekuatan kausalitas ini hampir sama dengan kausalitas searah antara trafik Fakultas Bisnis dan Ekonomika (M) terhadap trafik internet Perpustakaan (S).

Tabel 3. Uji Kausalitas

Kausalitas	<i>F-test</i>	<i>p-value</i>
$M \rightarrow E$	0.6556	0.4186
$S \rightarrow E$	2.1732	0.1429
$N \rightarrow E$	0.6993	0.4035
$E \rightarrow M$	$4.07 \times 10^{-5}$	0.9949
$S \rightarrow M$	3.4246	0.0650
$N \rightarrow M^*$	16.0850	$7.36 \times 10^{-5}$
$E \rightarrow S^*$	23.3597	$1.99 \times 10^{-6}$
$M \rightarrow S^*$	17.3228	$3.94 \times 10^{-5}$
$N \rightarrow S^*$	4.9041	$2.74 \times 10^{-2}$
$E \rightarrow N$	0.4377	0.5086
$M \rightarrow N$	2.4226	0.1205
$S \rightarrow N$	1.1438	0.2856

Analisis kausalitas trafik ini dapat dikaitkan dengan aktivitas pengguna internet di masing-masing lokasi trafik. Karakteristik pengguna internet di 3 lokasi yaitu Fakultas Teknik (E), Fakultas Bisnis dan Ekonomika (M) dan Kampus Ubaya Ngagel (N) mempunyai pengguna internet yang relatif tetap jumlahnya baik pengguna dari karyawan, dosen maupun mahasiswa. Sedangkan karakteristik pengguna internet di Perpustakaan cenderung berubah karena akses internet di lokasi ini diperuntukkan bagi mahasiswa menggunakan akses WiFi. Hubungan kausalitas trafik internet dan karakteristik pengguna diduga terdapat kaitan, namun masih perlu ada analisis lebih lanjut dan dalam.

#### 4. KESIMPULAN

Pemodelan trafik internet spasial yang melibatkan 4 lokasi pengukuran trafik secara agregat selama periode 1 tahun yaitu Fakultas Bisnis dan Ekonomika (M), Fakultas Teknik (E), Perpustakaan (S) dan Kampus Ubaya Ngagel (N) telah menghasilkan model VAR(1) yang sesuai dengan terlebih dahulu dilakukan validasi qqplot dan uji distribusi normal dari residu.

Analisis trafik dilakukan dengan menggunakan metoda Granger Causality pada model VAR dengan nilai  $p = 1$ , hasilnya menunjukkan bahwa trafik internet di Perpustakaan (S) dipengaruhi oleh trafik internet dari Fakultas Teknik (E), Fakultas Bisnis dan Ekonomika (M) dan Kampus Ubaya Ngagel (N). Trafik dari Fakultas Teknik (E) mempunyai pengaruh yang kuat dibandingkan yang lainnya. Diduga terdapat kaitan antara kausalitas trafik internet dengan karakteristik pengguna di masing-masing lokasi.

#### UCAPAN TERIMAKASIH

Makalah ini merupakan bagian dari hasil penelitian yang didukung secara penuh oleh LPPM Universitas Surabaya. Untuk itu penulis mengucapkan banyak terimakasih atas dukungan dana dari LPPM Universitas Surabaya.

#### DAFTAR REFERENSI

- [1] Dong-Yan Zhang, Ming-Zeng Hu, Hong-Li Zhang, Ting-Biao Kang (2005), "A Network Traffic Model Based On Measurement", Proc of 4<sup>th</sup> International Conferences on Machine Learning and Cybernetics, Guangzhou, China, 18-21 August 2005
- [2] H.Zare Moayedi, M.A. Masnadi-Shirazi (2008) ,"ARIMA Model for Network Traffic Prediction and Anomaly Detection", IT Symposium, Kuala Lumpur, Malaysia, 26-28 August 2008

- [3] Huda M. A. El Hag, Sami M. Sharif (2007), "An Adjusted ARIMA Model for Internet Traffic", AFRICON 2007
- [4] Kensuke Fukuda, Hideki Takayasu, Misako Takayasu(1999), "Spatial and Temporal Behavior of Congestion in Internet Traffic," Fractal Vol. 7 No.1 (1999) page 23-31
- [5] Luciano Rodrigues, Paulo Roberto Guardieiro (2004), "A Spatial and Temporal Analysis of Internet Aggregate Traffic at the Flow Level", IEEE Communications Society
- [6] A.K. Seth. (2005), "Causalities connectivity of evolved neural networks during behavior", Network: Computation in Neural Network Systems
- [7] Xiufang Du, Xiaofei Yan (2008), "Cointegration Analysis on The Relationship of China's Industrial Structure Change and The Oil Demand", 4<sup>th</sup> International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM '08.
- [8] Younghun Kim, Rahul Balani, Han Zhao, Mani B. Srivastava (2010), "Granger Causality Analysis on IP Traffic and Circuit-Level Energy Monitoring", Proceeding of The 2<sup>nd</sup> ACM Workshop on Embedded Sensing System for Energy Efficiency in Building, New York, USA
- [9] B. Pfaff (2008), "Analysis of Integrated and Cointegrated Time Series with R", Springer, Second Edition