

POLA PENGGUNAAN LALU-LINTAS INTERNET
PADA WAKTU TERTENTU MENGGUNAKAN
METODE PENAKSIR KERNEL

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk menempuh

Ujian Sidang Sarjana Matematika

Oleh:

MIFTAH ANDRIANSYAH

0396010296

Pembimbing:

Dr. Djati Kerami

Dra. Yekti Widyaningsih., MSi.



Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Indonesia

2003

LEMBAR PERSETUJUAN

Tugas akhir berjudul

POLA PENGGUNAAN LALU-LINTAS INTERNET PADA WAKTU TERTENTU
MENGUNAKAN METODE PENAKSIR KERNEL

Telah disetujui dan telah memenuhi syarat untuk diajukan sebagai salah satu kelengkapan dalam ujian akhir sarjana di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia.

Menyetujui

Depok, Januari 2003

Pembimbing

Pembimbing

Dr. Djati Kerami

Dra. Yekti Widyaningsih, MSi.

ABSTRAK

Miftah Andriansyah

NPM 0396010296

xi + 61 hal (2003)

Bibl. 6 (1990 - 2001)

Jumlah pemakai dan fungsi Internet sebagai alat komunikasi semakin meningkat. Oleh sebab itu, semakin menarik dan penting untuk diketahui. Diduga, pola penggunaan lalu-lintas Internet berbeda-beda.

Penulisan ini bertujuan untuk melihat pola fungsi distribusi penggunaan lalu-lintas Internet pada interval waktu (pukul 08.00-17.00 WIB) tertentu pada hari penggunaan yang berbeda-beda.

Metode kernel digunakan sebagai alat untuk melihat pola penggunaan lalu-lintas pada interval waktu tersebut.

Data yang digunakan diambil selama bulan September 2001, dan pengolahan data dilakukan dengan paket software R.

Hasil yang diperoleh, ternyata tidak ada pola seragam pada tanggal-tanggal pengambilan data.

Kata Kunci: metode kernel, pola penggunaan, lalu-lintas Internet.

DAFTAR ISI

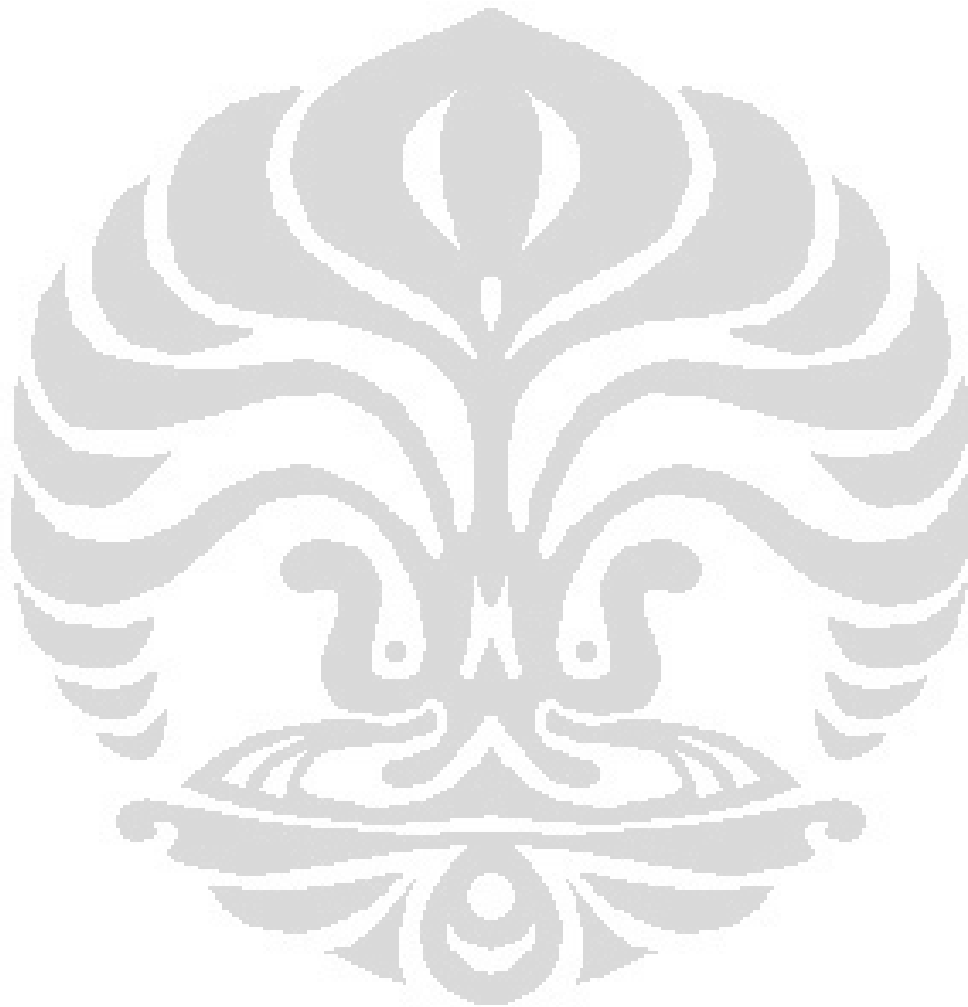
	Halaman
LEMBAR PERSETUJUAN	i
ABSTRAK	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	viii
KATA PENGANTAR	ix
BAB I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penulisan	2
1.3 Pembatasan Masalah	2
1.4 Metodologi	3
1.5 Organisasi Penulisan	4
BAB II. TINJAUAN PUSTAKA	6
2.1 Fungsi Distribusi	6
2.2 Internet	12
2.3 Komunikasi Data	14

BAB III. PENAKSIR METODE KERNEL	21
3.1 Motivasi Penggunaan Metode Kernel	21
3.2 Metode Kernel	22
3.3 Taksiran Parameter dan Uji Statistik	27
BAB IV. APLIKASI METODE KERNEL UNTUK PENAKSIRAN POLA PENGUNAAN LALU-LINTAS INTERNET	34
4.1 Data Penggunaan Internet	34
4.2 Pengolahan Data	37
4.3 Analisa Data	38
BAB V. PENUTUP	53
5.1 Kesimpulan	53
5.2 Saran	54
DAFTAR PUSTAKA	55
LAMPIRAN	56

DAFTAR GAMBAR

	Halaman
Gambar 2.1. Cara menjelaskan distribusi data; grafik fungsi densitas, grafik fungsi distribusi kumulatif, garis linear pada kertas probabilitas (metode grafik), rumus fungsi densitas	9
Gambar 2.2 Diagram umum jaringan Internet	13
Gambar 2.3 OSI Layer	14
Gambar 2.4 IP Data Diagram	16
Gambar 2.5 Pendekatan Diagram	17
Gambar 2.6 Pendekatan Sirkuit Virtual	18
Gambar 4.1. Tampilan informasi paket dalam <i>Ethereal</i>	36
Gambar 4.2 Tampilan protocol summary dalam <i>Ethereal</i>	37
Gambar 4.3.a Taksiran Densitas Penggunaan Lalu-lintas Internet pada Tanggal 12 September 2002	39
Gambar 4.3.b Taksiran Densitas Penggunaan Lalulintas Internet pada Tanggal 14 September 2002	40
Gambar 4.3.c Taksiran Densitas Penggunaan Lalulintas Internet pada Tanggal 17 September 2002	41
Gambar 4.3.d Taksiran Densitas Penggunaan Lalulintas Internet pada Tanggal 24 September 2002	42

Gambar 4.3.e Taksiran Densitas Penggunaan Lalulintas Internet
pada Tanggal 12 September 2002 43



DAFTAR TABEL

	Halaman
Tabel 3.1 Jenis Fungsi Kernel	22
Tabel 4.1 Score <i>NLL</i> Data Tanggal 12 September 2002	46
Tabel 4.2 Score <i>NLL</i> Data Tanggal 14 September 2002	47
Tabel 4.3 Score <i>NLL</i> Data Tanggal 17 September 2002	49
Tabel 4.4 Score <i>NLL</i> Data Tanggal 24 September 2002	50
Tabel 4.5 Score <i>NLL</i> Data Tanggal 27 September 2002	51

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Metode <i>Sheater & Jones (1991)</i> untuk pemilihan bandwidth h untuk penaksir kernel di \mathbb{R}	56
Lampiran 2 Metode <i>Unbiased Cross Validation</i> untuk pemilihan bandwidth h untuk penaksir kernel di \mathbb{R}	57
Lampiran 3 Metode <i>Rule of Thumb</i> untuk pemilihan bandwidth h untuk penaksir kernel di \mathbb{R}	58
Lampiran 4 Algoritma penaksir densitas kernel untuk ukuran bandwidth h sama dan fungsi kernel yang berbeda.....	59
Lampiran 5 Algoritma penaksir densitas kernel untuk ukuran bandwidth h tertentu dan fungsi kernel	60

KATA PENGANTAR

Bismillahirrahmaanirrahiim.

Alhamdulillahirrabbi'lalamiin, tiada kata yang patut penulis ucapkan selain puji syukur yang tidak terhingga kehadirat Allah SWT, atas segala rahmat, perlindungan dan petunjuk yang telah diberikan sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik.

Tugas akhir ini disusun untuk melengkapi persyaratan menempuh ujian sarjana di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia.

Ucapan terima kasih yang sebesar-besarnya penulis haturkan kepada yang terhormat:

1. Bapak **Drs. Suryadi SIS, MSc.** selaku Ketua Jurusan Matematika FMIPA Universitas Indonesia.
2. Bapak **Dr. Djati Kerami** dan Ibu **Dra. Yekti Widyaningsih, MSi.** yang telah memberikan bimbingan, saran, dorongan, dukungan, bantuan dan terlebih pengertian yang tulus dan tinggi serta penuh kesabaran dalam penulisan tugas akhir ini.
3. Segenap Dosen Jurusan Matematika FMIPA UI yang tidak dapat disebutkan satu persatu, yang membimbing dan mendidik penulis semasa kuliah.

4. Bapak **Prof. Suryadi HS** selaku Pembantu Rektor II Universitas Gunadarma yang memberikan kesempatan, waktu dan tempat untuk pengambilan data.
5. Bapak **Dr. Ing Adang Suhendra, MSc.** selaku dosen dan staff Universitas Gunadarma yang membantu dan mengasistensi dengan kesabaran selama penulisan ini.
6. Seluruh karyawan Jurusan Matematika FMIPA UI.
7. Rekan-rekan mahasiswa Jurusan Matematika FMIPA UI, khususnya angkatan 1996, terlebih buat Rahmi Rusin.
8. Mereka yang tidak dapat disebutkan namanya satu persatu yang telah memberikan dukungan, semangat dan waktunya kepada penulis.

Dengan penuh kerendahan hati dan setulus-tulusnya penulis ingin membaktikan semua ini kepada orang yang penulis hormati dan sayangi, Ibunda (Alm) dan Ayahanda, penulis sebagai jawaban untuk do'a yang tidak pernah putus, bimbingan, nasehat dan semangat demi keberhasilan dan keselamatan dalam segala penghidupan dan kehidupan penulis.

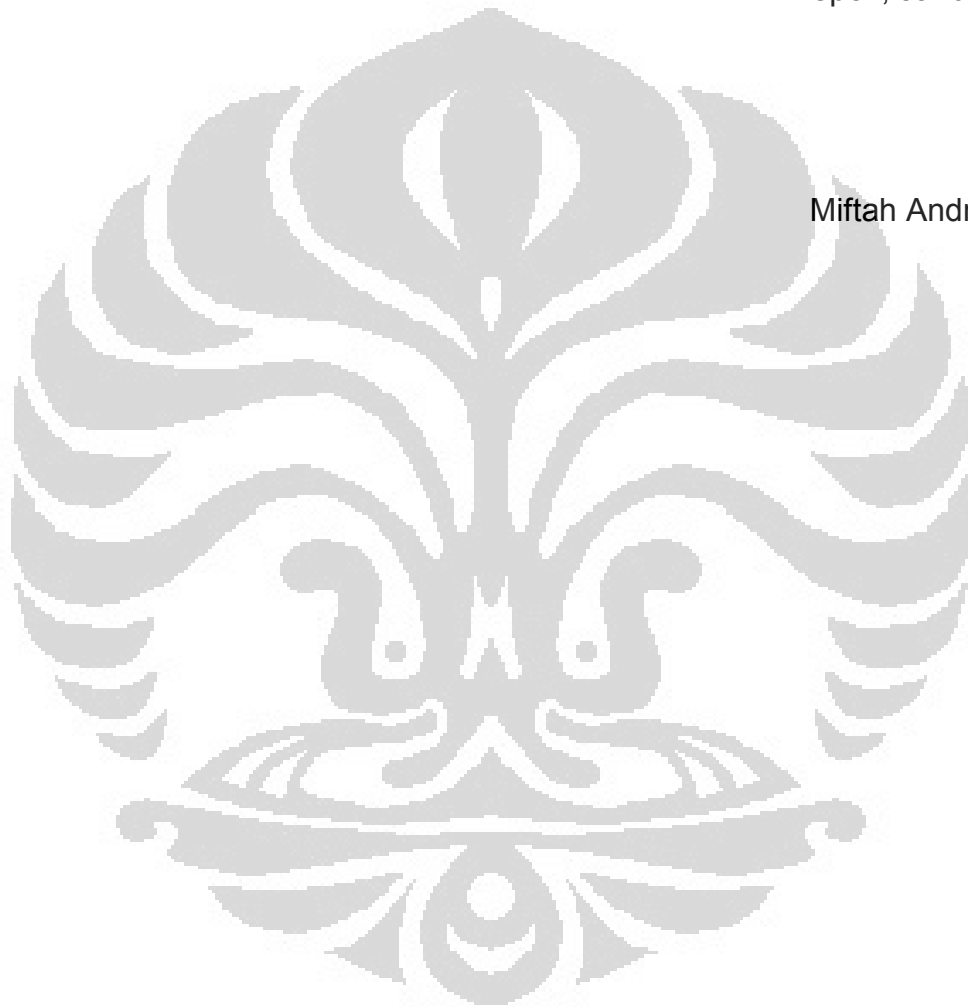
Penulis juga ucapkan terima kasih kepada kakak-kakak (Teh Nuke, A Adang, A Alif, dan The Elin), nenek, tante, dan keponakkan serta sanak saudara yang telah memberikan semangat dan dukungannya. Semoga Allah senantiasa melimpahkan rahmat dan hidayah-Nya kepada kita semua. Amiin.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna, untuk itu saran, kritik, maupun masukan lain yang bermanfaat bagi

kesempurnaan tugas akhir ini sangat penulis harapkan. Akhir kata, penulis berharap semoga tugas akhir ini dapat memberikan sumbangan yang berarti terhadap perkembangan pemikiran dan ilmu pengetahuan.

Depok, Januari 2003

Miftah Andriansyah



BAB I

PENDAHULUAN

1.1 Latar Belakang

Penggunaan Internet sebagai alat komunikasi dan pertukaran informasi secara digital terus mengalami peningkatan dari tahun ke tahun. Diperkirakan pengguna Internet di seluruh dunia mencapai angka 612 juta pengguna yang tersambung ke Internet pada bulan September 2002 dan kemungkinan jumlahnya akan terus bertambah pada tahun berikutnya.

Karena meningkatnya penggunaan Internet untuk komunikasi data, diperlukan pengaturan supaya proses komunikasi data berjalan dengan lebih baik. Untuk melakukan pengaturan penggunaan Internet diperlukan pengetahuan mengenai karakteristik penggunaan tersebut.

Kepentingan dan manfaat dari penggunaan lalu-lintas Internet, bermacam-macam untuk surat-menyurat, toko on-line, penyedia jasa hiburan misal audio-video, pendidikan dan penelitian. Data (atau selanjutnya disebut paket) yang ditransmisikan juga bermacam dari yang kecil (format teks) hingga besar seperti audio dan video streaming.

Penggunaan Internet yang baik adalah manakala adanya efisiensi antara penggunaan data yang ditransmisikan dalam bandwidth yang tersedia. Salah satu contoh permasalahan yang berkaitan dengan ketidakefisienan

penggunaan Internet sebagai contoh adalah pengiriman paket file video berukuran 5GB pada 19.2Kb/s bandwidth atau pengiriman teks pada situs jaringan pada 100Mb/s bandwidth. Pengetahuan mengenai bandwidth dan penggunaan aplikasi Internet kiranya dapat menghindarkan dari ketidakefisienan tersebut. Hal ini dapat diatasi dengan cara, mengadaptasi ukuran paket yang dikirim atau memilih webserver yang memiliki bandwidth lebih besar. Adaptasi ukuran paket yang dikirim atau memilih webserver yang memiliki bandwidth lebih besar, salah satunya memeriksa ukuran paket dalam protokol Hypertext Transfer Protocol (HTTP), Transmission Control Protocol (TCP), Simple Mail Transfer Protocol (SMTP) dan Internet Control Message Protocol (ICMP) serta besarnya bandwidth dimiliki jaringan Internet. Pemeriksaan pada protokol HTTP, TCP, SMTP, ICMP didasarkan karena penggunaannya yang cukup dominan dalam pengiriman paket. Pemeriksaan bandwidth cukup dengan mengetahui besarnya bandwidth suatu jaringan.

1.2 Tujuan

Mengetahui pola penggunaan lalu-lintas Internet pada interval waktu (jam) tertentu dengan melihat fungsi densitas penggunaan.

1.3 Pembatasan Masalah

Dalam penulisan ini menggunakan metode Kernel sebagai alat untuk menaksir fungsi densitas banyaknya paket yang melewati titik gerbang atau gateway jaringan Internet. Langkah-langkah dalam melakukan metode

kernel: Pertama, melihat grafik kurva yang dihasilkan dan mencari kandidat fungsi distribusi dari kurva yang mendekati kurva hasil plot kurva pertama. Kedua, melakukan uji statistik untuk kandidat-kandidat fungsi distribusi tersebut. Ketiga, melakukan penaksiran parameter-parameter fungsi distribusi. Keempat, mengambil keputusan taksiran distribusi yang paling mendekati. Kelima, melakukan analisa terhadap taksiran fungsi distribusi pengaruhnya terhadap kinerja *traffic* Internet.

Dalam metode kernel dibutuhkan dua parameter yaitu: parameter pemulusan h , dengan h terbaik (BAB II), dan fungsi kernel, dan fungsi kernel Gaussian.

Data lalu-lintas penggunaan Internet adalah total jumlah paket yang melewati titik gerbang jaringan (router). Dipilih secara acak selama bulan September 2002 didapat tanggal 11, 12, 14, 16, 17, 18, 19, 24, dan 27 September 2002. Untuk tiap-tiap hari "penangkapan" paket (package capturing), waktu yang dipilih adalah mulai pukul 08.00-17.00 WIB. Paket diambil (capture) pada router jaringan komputer Universitas Gunadarma.

1.4 Metodologi

Metodologi penulisan tugas akhir ini, yaitu: Data diperoleh dengan "menangkap" paket-paket yang melewati router dengan menggunakan software Ethereal. Kemudian dihitung besarnya paket dalam protokol HTTP, SMTP, ICMP, dan TCP. Kemudian dijumlahkan besarnya paket dalam protokol-protokol tersebut untuk tiap-tiap jam, dimulai pukul 08.00-17.00. Sebagai contoh, data lihat pada lampiran data. Langkah selanjutnya yaitu

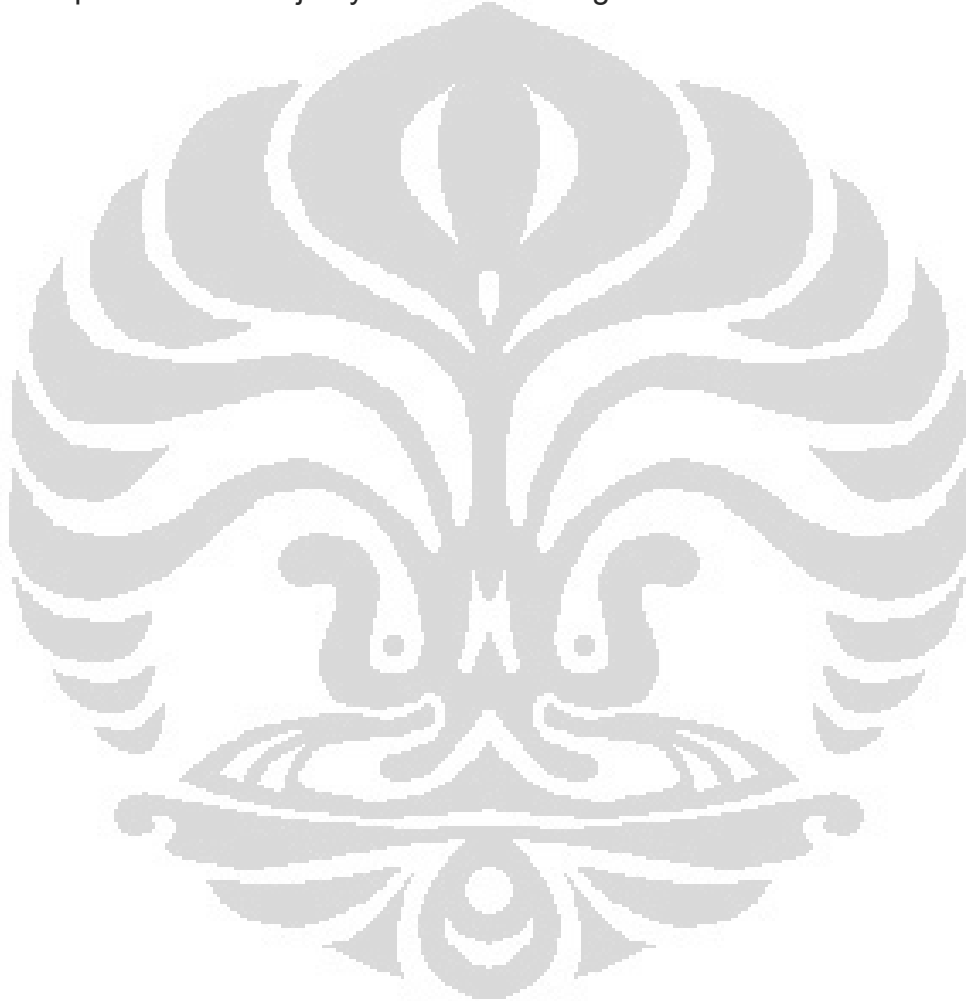
melihat pola penggunaan lalu-lintas Internet dengan membuat kurvanya untuk data pada tanggal-tanggal 12, 14, 17, 24, dan 27 September 2002 menggunakan metode penaksir densitas kernel. Pembuatan kurva masing-masing data menggunakan aplikasi/software R versi 1.4.1. Dari kurva-kurva tersebut, ditentukan jenis taksiran fungsi densitasnya. Hasil taksiran fungsi densitas yang diketahui, akan dilihat apakah ada kecenderungan yang sama fungsi densitas pola penggunaan lalu-lintas Internet yang ditentukan.

1.5 Organisasi Penulisan

Organisasi penulisan dalam tugas akhir ini meliputi:

- Bab I, Pendahuluan yang membahas mengenai latar belakang penulisan, tujuan penulisan, pembatasan masalah, metodologi dan organisasi penulisan.
- Bab II, Tinjauan Pustaka yaitu pengetahuan tentang Internet dan komunikasi data, pengiriman paket dari sumber ke tujuan, jenis protokol yang digunakan dalam penulisan ini, dan tentang Ethernal. Tinjauan pustaka ini juga membahas tentang konsep dari fungsi distribusi dan jenis-jenisnya.
- Bab III, yaitu penaksir densitas kernel, metode kernel, metode pemilihan parameter pemulusan h , penaksiran parameter fungsi densitas, uji *Goodness of Fit*.

- Bab IV, Aplikasi Metode Kernel dan Analisa Data yang terkait dengan proses pengambilan data, pengolahan data, nilai taksiran parameter menggunakan software FIT dan analisa hasil taksiran fungsi densitas.
- BAB V, Kesimpulan dan Saran, untuk perbaikan dan pengembangan penelitian selanjutnya di masa datang.



BAB II

TINJAUAN PUSTAKA: FUNGSI DISTRIBUSI, INTERNET DAN KOMUNIKASI DATA

Dalam bab ini akan membahas definisi dari fungsi distribusi,, dan definisi dari paket (package) yang ditransmisikan dalam Internet dari satu titik ke titik lainnya di jaringan.

2.1 Fungsi Distribusi

Fungsi distribusi adalah suatu konsep dasar dalam statistika yang digunakan sebagai penentu besarnya probabilitas untuk suatu selang yang diberikan, yang dituliskan dengan rumus berikut:

$$P(a < X < b) = \int_a^b f(x) dx$$

dimana $f(x)$ fungsi densitas dari peubah acak X .

Pada umumnya fungsi densitas suatu variabel random tidak dapat diperoleh secara langsung. Misalkan kita memiliki sekumpulan data observasi yang diasumsikan sebagai sampel dari fungsi densitas yang tidak diketahui. Untuk itu dilakukan suatu pengukuran atau disebut sebagai penaksiran fungsi densitas untuk mendapatkan fungsi densitas yang mewakili suatu populasi. Penaksiran fungsi densitas dapat dilakukan dengan menggunakan beberapa

metode, diantaranya: Histogram, Penaksir Naive, Penaksir Kernel, metode K-Nearestneighbour, metode Variabel Kernel, Penaksir Deret Othogonal, Penaksir *Maximum Penalized Likelihood*, Fungsi bobot umum.

Ada tiga metode yang biasa digunakan untuk menaksir fungsi densitas variabel random, yaitu:

- Metode Histogram
- Metode Grafik
- Metode Kernel

Pendekatan dalam menaksir fungsi densitas dapat dilakukan secara parametrik dan non-parametrik.

- *Pendekatan parametik*, dimana kita akan menaksir nilai-nilai parameter untuk fungsi densitas yang telah diketahui. Dengan kata lain, menaksir fungsi densitas ekivalen dengan menaksir nilai parameter. Sebagai contoh, kita sudah mengetahui bahwa variabel acak X berdistribusi normal dengan nilai mean μ dan variansi σ^2 yang tidak diketahui, yakni,

$$f = f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right)$$

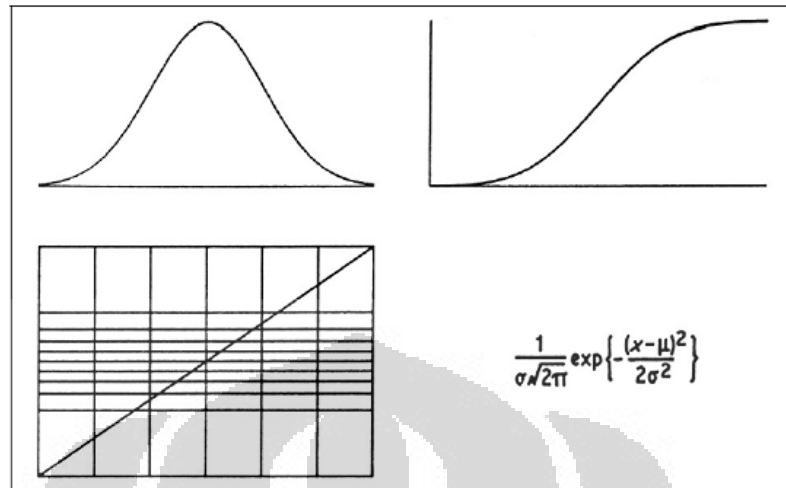
Maka terlihat jelas taksiran densitas X adalah fungsi densitas normal dengan μ dan σ adalah parameternya.

- *Pendekatan nonparametrik*, yaitu pendekatan untuk menaksir kurva densitas dari variabel acak. Pendekatan ini masuk akal pada saat kita

tidak mempunyai informasi yang cukup mengenai bentuk dan kelas dari densitas sesungguhnya. Data diberi peluang untuk "berbicara mengenai sifat atau perilakunya sendiri" Dapat pula diperhatikan, Berapa banyak modus dari densitas? Dimana titik-titik pengelompokkan data?

Pendekatan yang digunakan dalam penelitian ini dalam rangka untuk mengetahui kepadatan penggunaan lalu-lintas Internet dalam penulisan ini adalah *pendekatan nonparametrik*.

Karena fungsinya, penaksiran fungsi densitas sangat ideal untuk menampilkan atau mempresentasikan data untuk memberikan penjelasan dan ilustrasi atas kesimpulan yang diambil. Ada beberapa cara yang digunakan untuk menjelaskan distribusi suatu data (misal untuk fungsi distribusi normal) yaitu: menggambar grafik fungsi densitasnya, menggambar grafik fungsi distribusi kumulatif, garis linear pada kertas probabilitas (metode grafik) dan menuliskan formula fungsi densitasnya. Secara visual lihat gambar 2.1.



Gambar 2.1. Cara menjelaskan distribusi data; grafik fungsi densitas, grafik fungsi distribusi kumulatif, garis linear pada kertas probabilitas (metode grafik), rumus fungsi densitas

Secara umum, tujuan penaksiran fungsi densitas suatu data adalah untuk mendapatkan kurva fungsi densitas yang merupakan kurva mulus dengan variansi sampling tidak besar dan tidak hilangnya informasi penting dari suatu data.

2.1.1 Penaksir Fungsi Distribusi Data Non-Parametrik

Dari beberapa metode penaksir fungsi distribusi suatu data, tiga metode yang sering digunakan yaitu:

1. Metode Histogram

Metode histogram adalah penaksir fungsi densitas yang tertua dan paling sering digunakan. Diberikan suatu titik awal x_0 dan lebar *bin* h , didefinisikan *bin* suatu histogram adalah interval $[x_0 + mh, x_0 + (m + 1)h]$

untuk m anggota bilangan bulat.

Definisi histogram,

$$f(x) = \frac{1}{nh} \text{ (banyaknya } Xi \text{ pada bins yang sama sebagai } x \text{)}$$

Dengan menganggap lebar *bin* yang bervariasi, histogram dapat digeneralisir dan didefinisikan sebagai:

$$f(x) = \frac{1}{n} \times \frac{\text{(banyaknya } Xi \text{ dalam bin yang sama sebagai } x \text{)}}{\text{(lebar bin yang mengandung } x \text{)}}$$

Dalam pembentukan histogram, yang harus diperhatikan yaitu penentuan titik awal x_0 dan pemilihan lebar *bin* h . Pemilihan *bin* h dapat menentukan kelulusan dari histogram. Berikut adalah langkah-langkah yang diambil secara umum dalam melakukan perhitungan histogram sebagai penaksir fungsi densitas:

- Membagi garis bilangan real menjadi bin-bin

$$B_j = [x_0 + (j - 1)h, x_0 + jh), j \in \mathbb{Z}, \mathbb{Z} = \text{bilangan bulat.}$$

dimana

$h = \text{lebar bin histogram, } h > 0, x_0 = \text{titik awal histogram.}$

— Menghitung jumlah data yang terletak pada setiap *bin*.

Sehingga taksiran fungsi densitas secara umum ditulis sebagai:

$$\hat{f}_h(x) = (nh)^{-1} \sum_{i=1}^n \sum_{j \in \mathbb{Z}} I(X_i \in B_j) I(x \in B_j)$$

dengan $I = \text{fungsi indikator}$, dimana $j \in \mathbb{Z}$ dan $i = 1, 2, 3, \dots, n, I(x \in B_j) = 1$ untuk $x \in B_j$.

Dalam penggambaran fungsi histogram, $\hat{f}_h(x)$ menunjukkan tinggi batang-batang (bar) histogram untuk setiap *bin* dan luas dari seluruh batang histogram sama dengan satu. Namun dalam penerapannya, histogram kurang begitu disukai oleh pengguna, karena diskontinuitas yang dihasilkan oleh histogram sehingga menyebabkan kesulitan yang signifikan dalam diferensiasi atau penurunan fungsi densitas yang dibutuhkan. Histogram kurang efisien jika diterapkan untuk menaksir fungsi densitas pada analisis cluster dan analisis diskriminan non-parametrik.

Penerapan histogram cukup efisien untuk presentasi dan eksplorasi data. Oleh karena itu dibutuhkan metode yang lebih efisien untuk

menaksir fungsi densitas yang tidak bergantung pada pemilihan titik awal.

2. Metode Grafik

Metode Grafik adalah penaksir fungsi densitas suatu variabel random yang membandingkan kandidat fungsi densitas dengan logaritma fungsinya sebagai suatu hubungan linear. Semakin mendekati kelinearan, maka taksiran fungsi densitas mendekati sama.

3. Metode Kernel

Tidak seperti metode penaksiran fungsi densitas sebelumnya (histogram dan grafik), metode kernel adalah metode penaksiran fungsi distribusi yang memberikan fungsi yang kontinu.

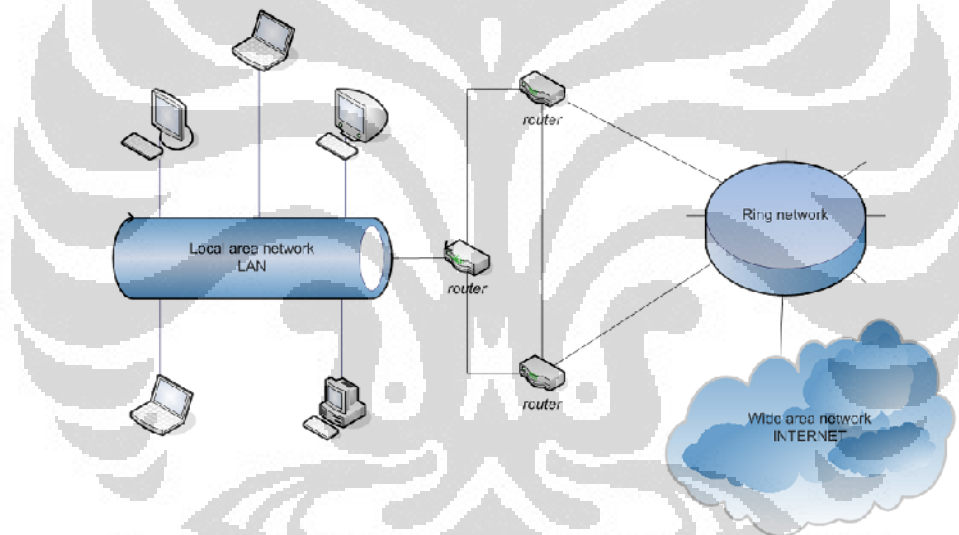
2.2 Internet

Internet adalah media komunikasi yang menghubungkan komputer satu dengan lainnya dalam suatu sistem jaringan. Data yang dikomunikasikan, yaitu dalam bentuk paket-paket yang mempunyai standarisasi dan ukuran tertentu. Dan pengertian Internet menurut The Federal Networking Council/FNC yang dikutip dari Laporan Bulanan Internet, Oktober 1995, yaitu suatu sistem informasi global, yang:

1. menghubungkan tiap komputer-komputer yang masing-masing memiliki alamat yang unik.

2. untuk mendukung komunikasi dengan menggunakan rangkaian Transmission Control Protocol/Internet Protocol (TCP/IP) atau sub-barisan dari Internet Protocol IP), dan protokol lainnya yang bersesuaian (compatible) dengan IP
3. menyediakan, menggunakan atau membuatnya dapat diakses oleh umum maupun pribadi dalam Internet.

Berikut adalah ilustrasi umum mengenai jaringan Internet yang digambarkan pada gambar 2.2.

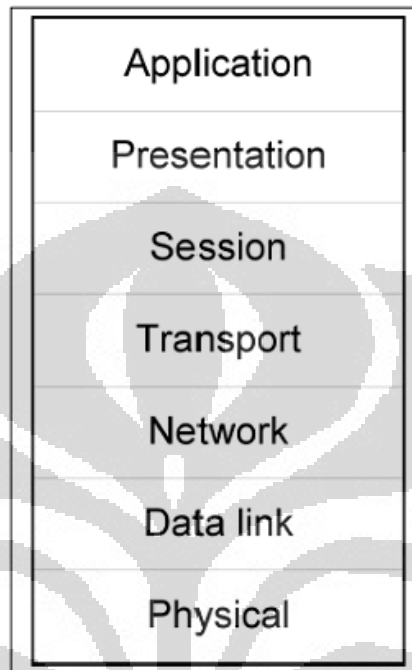


Gambar 2.2 Diagram umum jaringan Internet

2.2.1 Lapisan-lapisan Internet

Pengirim data pada jaringan komputer mengubah data yang berasal dari aplikasi seperti teks, suara dan lain-lain ke dalam bentuk data fisik biner 0 atau 1. Perubahan data tersebut diproses melalui beberapa tahapan,

dimana standar tahapan perubahan data adalah lapisan model *Open System Interconnection* (OSI) yang dijelaskan pada gambar 2.3.



Gambar 2.3 OSI Layer

2.3 Komunikasi Data

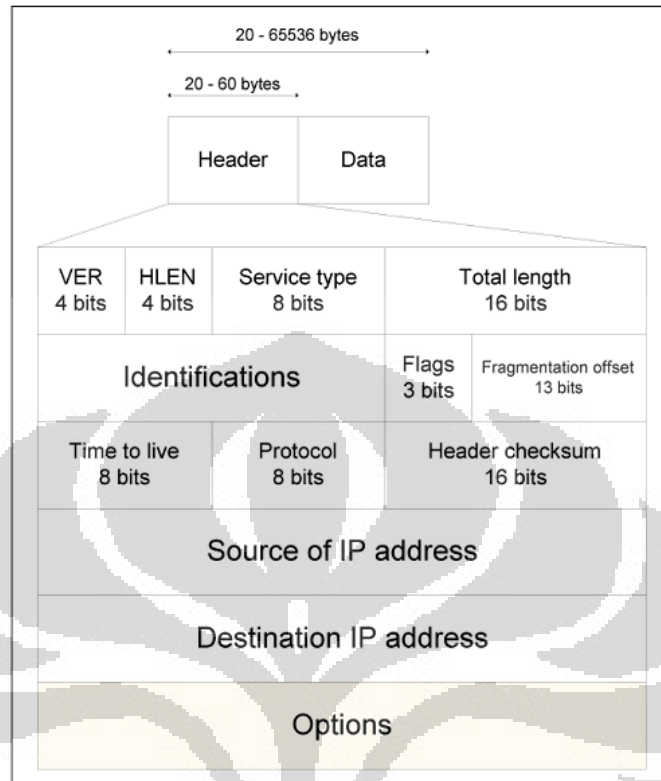
Paket data Internet yang ditransfer dari komputer satu ke komputer lainnya merupakan inti dari komunikasi Internet dimana mengandung berbagai macam file yang coba dipertukarkan melalui Internet (dengan standar yang telah ditetapkan). Paket data Internet itu sendiri saling mempengaruhi terhadap unsur-unsur Internet lainnya, diantara: lebar bandwidth, ketepatan, kecepatan, kesempurnaan pengiriman paket tersebut dan hal-hal lainnya. Jumlah atau banyaknya paket data Internet adalah variabel random yang dapat menjelaskan banyak hal, diantaranya kualitas

layanan Internet. Seperti variabel random lainnya, maka banyaknya paket data Internet memiliki distribusi.

2.3.1 Pengertian Paket Internet

Data yang dikirim melalui Internet, terlebih dahulu dipecah menjadi bagian-bagian kecil dengan ukuran tertentu yang dinamakan paket. Satuan terkecil paket yaitu byte. Pengiriman paket melalui Internet melalui beberapa tahapan dari titik ke titik (*node to node*), paket tersebut kemudian disimpan untuk selanjutnya dikirim ke tujuannya menurut urutan dan arahnya, yang sesuai dengan informasi yang diberikan. Proses pengiriman dengan urutan dan arah tersebut dinamakan *routing*, yang diilustrasikan pada gambar 2.4.

Paket terdiri dari dua bagian, yaitu: *header* (kepala) dan data (badan). *Header* memiliki ukuran 20-60 byte yang berisikan informasi penting untuk pengiriman paket.



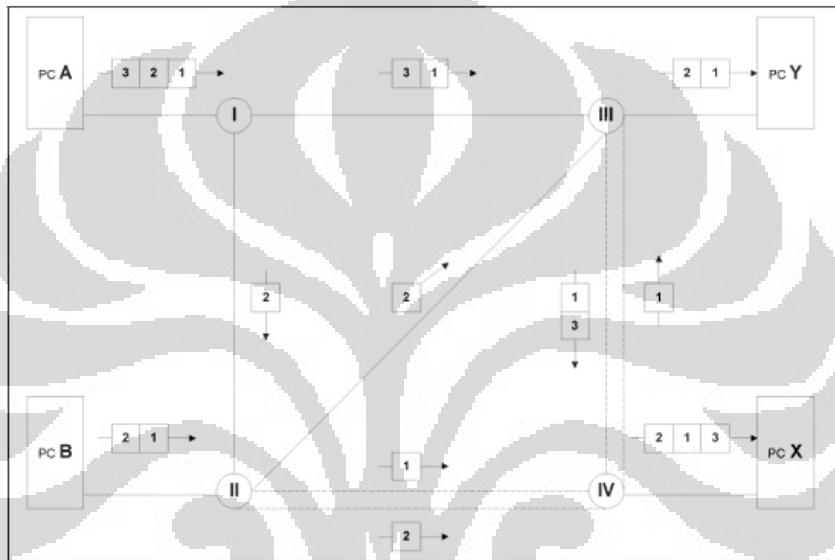
Gambar 2.4 IP Data diagram

2.3.2 Pengiriman paket Internet

Pengiriman paket dari komputer sumber ke tujuan dapat dicapai dengan dua pendekatan:

1. Datagram: Dalam pendekatan ini, setiap paket diperlakukan (saling bebas) independen terhadap paket lainnya. Misalkan paket dikirim dari komputer PC A dan PC B ke komputer PC X dan PC Y. PC A mengirim tiga paket ke PC X. Ada empat titik yang dapat dilalui oleh tiga paket tersebut. Paket nomor 3 dan nomor 1 dikirim melalui jalur I-III dan III-IV, sedangkan paket nomor 2 dikirim melalui jalur yang berbeda yaitu jalur I-II dan II-IV hingga sampai ke PC X. PC

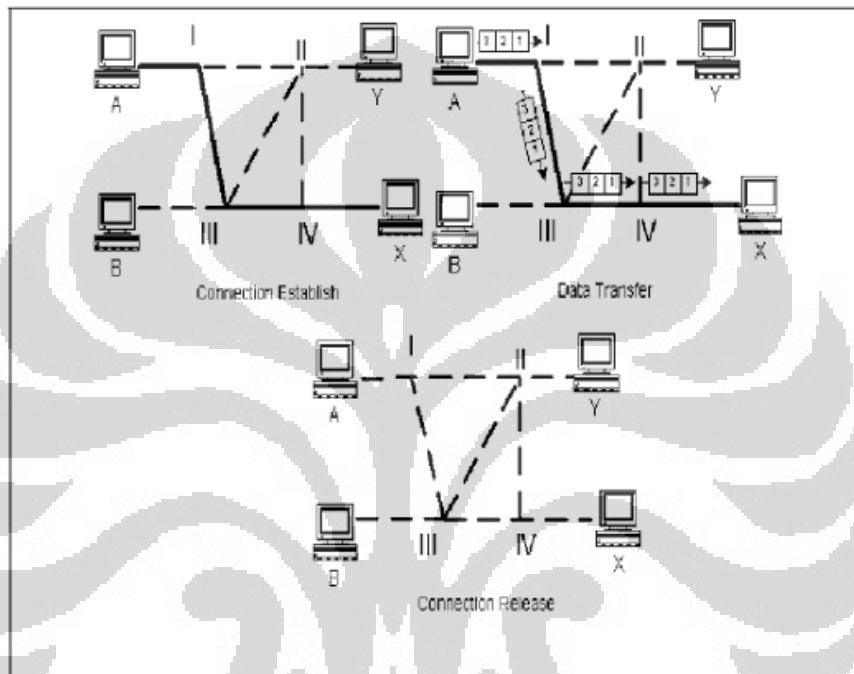
B mengirim dua paket, paket nomor 1 dikirim melalui jalur II-IV, dan paket nomor 2 dikirim melalui jalur II-III hingga sampai ke PC Y. Dapat dilihat pada gambar 2.5 bahwa pengiriman paket tidak selalu berurutan dan jalur yang ditempuh tidak tetap, berdasarkan padat atau tidaknya suatu jalur.



Gambar 2.5 Pendekatan Diagram

2. Sirkuit virtual: Dalam pendekatan ini, paket dikirim secara berurutan, dengan terlebih dahulu menetapkan jalur-jalur mana yang akan ditempuh oleh paket hingga tujuan. Misalkan paket akan dikirim dari komputer PC A ke komputer PC X. Selain PC A dan PC X terdapat komputer lainnya yaitu komputer PC B dan PC Y dalam satu jaringan. Langkah-langkahnya adalah, pertama ditentukan dahulu jalur yang akan dilalui oleh paket dengan menetapkan hubungan (connection establishment) didapat jalur I-II-IV, kedua

pengiriman paket sesuai dengan urutannya pada jalur tersebut, dan ketiga, setelah paket sampai ke tujuan, hubungan yang ditetapkan diputus (connection release). Pengiriman paket dengan pendekatan sirkuit virtual digambarkan pada gambar 2.6.



Gambar 2.6 Pendekatan Sirkuit Virtual

2.3.3 Protokol Komunikasi Internet

Pengertian protokol dalam Internet adalah sekumpulan peraturan yang mengatur komunikasi data. Protokol bertugas mendefinisikan apa yang dikomunikasikan, bagaimana komunikasi terjadi dan kapan komunikasi terjadi. Ada tiga elemen kunci dari protokol yaitu: sintak (syntax), semantik (semantic), waktu (timing).

- *Syntax* merujuk pada struktur atau format data yaitu urutan data yang akan ditampilkan. Sebagai contoh, suatu protokol sederhana terdiri dari 8 bit pertama data sebagai alamat pengirim, 8 bit kedua sebagai alamat penerima, dan sisanya adalah pesan (message).
- *Semantic* merujuk pada arti untuk bagian-bagian bit. Maksudnya, untuk suatu pola tertentu akan diterjemahkan dan diambil tindakan tertentu. Sebagai contoh, apakah suatu alamat menunjukkan rute yang akan dilalui atau menunjukkan tujuan akhir suatu pesan.
- *Timing* merujuk pada dua karakteristik: waktu pengiriman data dan seberapa cepat data dikirim. Sebagai contoh, jika pengirim menghasilkan data pada bandwidth 100 Megabits per second (Mbps) namun penerima hanya bisa memproses data pada bandwidth 1Mbps, transmisi akan kelebihan muatan (overload) pada penerima, dan data kemungkinan besar akan hilang.

Beberapa protokol dalam Internet, diantaranya yaitu:

- Hypertext Transfer Protocol (HTTP) adalah protokol yang utamanya digunakan untuk mengakses data pada WorldWideWeb (WWW). Protokol ini mentransfer data dalam bentuk plaintext (text-only), hypertext, audio, video dan lain sebagainya
- Transmission Control Protocol (TCP) adalah protokol yang menyediakan *service full transport*. Protokol ini, sebelum mengirim data, terlebih dahulu melakukan hubungan *handshaking* antara pengirim dan

penerima. Sebelum *handshaking* terjadi, tidak akan terjadi proses pengiriman data. Hal tersebut karena TCP memiliki karakteristik berorientasi koneksi (*connection oriented*).

- Simple Mail Transfer Protocol (SMTP) adalah protokol TCP yang mendukung pengiriman e-mail(*electronic-mail*). SMTP adalah sistem untuk mengirim pesan ke tiap komputer pengguna berdasarkan alamat e-mail masing-masing. SMTP mendukung:
 - Pengiriman satu pesan ke satu atau lebih penerima.
 - Pengiriman pesan yang mengandung teks, suara/*voice*, video, atau grafik.
 - Pengiriman pesan ke pengguna pada jaringan-jaringan di luar Internet (dari jaringan lokal ke jaringan global).
- Internet Control Message Protocol (ICMP) adalah protokol yang mengirim pesan atas kesalahan (*error*) yang terjadi pada pengiriman paket dan mengirim pesan mengenai informasi yang dibutuhkan dari *host* atau router.

Mekanisme pesan kesalahan (*error message*) terbagi menjadi dua yaitu pelaporan kesalahan dan koreksi kesalahan. Contoh pesan yang diberikan apabila router pengirim harus mengabaikan data karena tidak dapat mencapai router tujuan akhirnya.

BAB III

PENAKSIR METODE KERNEL

Dalam bab ini akan dibahas mengenai perangkat penaksir fungsi densitas pada tugas akhir ini, yaitu metode kernel.

3.1 Motivasi Penggunaan Metode Kernel

Ada beberapa pertimbangan yang mendasari penggunaan metode kernel untuk penulisan tugas akhir ini, diantaranya:

- Menurut *Kevin Lai dan Mary Baker* dalam penelitiannya menggunakan penaksir densitas kernel untuk melihat kepadatan hasil filterisasi terhadap sampel-sampel yang mempengaruhi kepadatan penggunaan lalu-lintas Internet dengan bandwidth tertentu.
- metode penaksir kernel dapat menyajikan penaksiran yang baik dibandingkan dengan metode histogram (selain pemilihan parameter pemulusan h , kesulitan lainnya yaitu pemilihan titik awal x_0 yang dapat memberikan interpretasi yang berbeda-beda), dan kelemahan histogram lainnya yaitu kediskontinuan pada batas-batas kelas.
- Ditinjau dari kecepatan perhitungan, metode kernel lebih banyak membutuhkan waktu dibandingkan dengan histogram. Namun keandalan dan keakuratan metode kernel lebih baik dibandingkan histogram.

3.2 Metode Kernel

Ada beberapa jenis fungsi kernel, antara lain:

Tabel 3.1 Jenis Fungsi Kernel

No	Kernel	Fungsi Kernel $K(u)$
1	Gaussian	$\frac{1}{\sqrt{2\pi}} \exp(-0.5u^2)$
2	Uniform	$\frac{1}{2} I(u \leq 1)$
3	Triangle	$(1 - u) I(u \leq 1)$
4	Quartic	$\frac{15}{16} (1 - u^2)^2 I(u \leq 1)$
5	Epanechnikov	$\frac{3}{4} (1 - u^2) I(u \leq 1)$
6	Triweight	$\frac{35}{32} (1 - u^2)^3 I(u \leq 1)$
7	Cosinus	$\frac{\pi}{4} \cos(\frac{\pi}{2}u) I(u \leq 1)$

Dan dalam penelitian tugas akhir ini, jenis fungsi kernel yang digunakan adalah Fungsi Kernel Gaussian.

3.2.1 Definisi Metode Kernel

Diasumsikan bahwa distribusi dari peubah acak X memiliki fungsi densitas f , maka

$$f(x) = \lim_{h \rightarrow 0} \frac{1}{2h} P(x - h < X < x + h)$$

Probabilitas $P(x - h < X < x + h)$ ditaksir dengan $f_i n$, dimana n adalah ukuran sampel dengan peubah acak X_1, X_2, \dots, X_n dan f_i adalah banyaknya

X_i yang ada dalam interval $(x - h, x + h)$. Sehingga suatu penaksir untuk f dapat ditentukan sebagai:

$$f(x) = \frac{1}{2h} [\text{banyak } X_i \text{ di } (x - h, x + h)].$$

pilih nilai h yang kecil. Fungsi \hat{f} disebut penaksir naif.

Jika $\hat{f}(x)$ dinyatakan dengan fungsi bobot w , maka

$$w(u) = \begin{cases} \frac{1}{2}, & \text{jika } |u| < 1 \\ 0, & \text{jika } |u| \geq 1 \end{cases}$$

Sehingga didapat persamaan:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n w\left(\frac{x - X_i}{h}\right)$$

Sama seperti penaksir histogram, penaksir naif menghasilkan taksiran fungsi \hat{f} yang tidak kontinu (fungsi tangga). Hal ini menyebabkan kesulitan dalam melakukan analisa dan pengukuran lebih lanjut. Namun apabila fungsi w digantikan dengan fungsi K yang memenuhi:

$$\int_{\mathbb{R}} K(t) dt = 1$$

dimana K simetris terhadap titik 0.

Untuk itu didapat penaksir kernel untuk suatu fungsi densitas f , dengan definisi lengkapnya sebagai berikut:

Penaksir $\hat{f} : R \rightarrow [0, \infty)$ dihitung pada basis nilai dari n-elemen banyak

Sampel acak x_1, x_2, \dots, x_n . Bentuk dasar penaksir kernel didefinisikan sebagai:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

dimana,

x = suatu nilai tertentu $\in R$

X_i = peubah acak independen dan berdistribusi identik

K = fungsi kernel

n = ukuran sampel

h = parameter pemulusan (lebar bandwidth), $h > 0$

Dimana fungsi kernel $K: R \rightarrow [0, \infty)$ memenuhi kondisi:

$$\int_{\mathbb{R}} K(t) dt = 1$$

dan untuk $x \in \mathbb{R}$

$$K(x) = K(-x)$$

$$K(0) = K(x)$$

Untuk memenuhi kriteria fungsi densitas, maka: $\int_{\mathbb{R}} \hat{f}_h dx = 1$

3.2.2 Pemilihan Kernel dan parameter pemulusan (bandwidth)

Pemilihan parameter pemulusan h merupakan hal yang penting dalam penaksiran densitas kernel. Perbedaan nilai parameter pemulusan h untuk suatu data akan memberikan tampilan kurva yang berbeda, yang kemungkinan dapat memberikan penafsiran yang berbeda pula. Untuk itu adalah penting untuk memilih parameter pemulusan h yang optimal untuk data penelitian ini.

Pemilihan parameter pemulusan yang optimal adalah parameter pemulusan h yang meminimumkan *Mean Square Error (MSE)* atau *Mean Integrated Square Error (MISE)*. Namun dalam formula matematika MSE dan MISE, masih mengandung fungsi f dan f'' yang tidak diketahui dan. Oleh karena itu, sampai saat ini, kita hanya dapat memperoleh h yang optimal untuk taksiran f , jika kita mengetahui fungsi f -nya.

Berikut adalah teknik yang pemilihan h jika kita tidak mengetahui fungsi f :

1. *Rule of Thumb (hrot)*

Adapun untuk nilai h -nya:

$$\hat{h}_0 = 1.06 \min \left(\sigma, \frac{\hat{R}}{1.34} = \pi r^2 \right) n^{-0.2}$$

2. *Oversmoothing (hos)*

Adapun untuk nilai h -nya:

$$\hat{h}_{0s} = 1.144 \hat{\sigma} n^{-0.2}$$

3. Validasi Silang (CVLs)

Validasi silang merupakan salah satu teknik untuk mencari h optimal untuk penaksir fungsi densitas. Teknik validasi silang yang digunakan dalam penulisan yaitu validasi silang maksimum likelihood (CVkl) dan validasi silang least square (CVls).

Karena penaksir kernel dijabarkan dari penaksir naif, maka penaksir kernel tidak tergantung pada pemilihan titik awal (seperti pada penaksir histogram yang menentukan titik awal). Seperti diketahui penaksir kernel mempunyai dua parameter yaitu fungsi kernel dan bandwidth h . Secara teoretis dapat dijelaskan bahwa dengan fungsi kernel yang berbeda akan menghasilkan bentuk global taksiran fungsi densitas sama. Hal tersebut tidak berlaku pada nilai parameter h yang akan mempengaruhi bentuk global taksiran fungsi densitas untuk h yang berbeda. Oleh karena itu penentuan nilai h yang tepat menjadi hal yang penting untuk dilakukan. Validasi silang dapat dilakukan untuk menentukan nilai h yang optimal. Ada dua jenis validasi silang, yaitu: Validasi silang *Kulbak Liebler* dan validasi silang kuadrat terkecil. Yang akan dibahas dalam penulisan ini adalah validasi silang *Kulbak Liebler*.

- Validasi silang Kulbak Liebler (KL) yaitu validasi silang yang memaksimumkan fungsi probabilitas.

Mula-mula akan diuji:

$$H_0 \quad \hat{f}_h(x) = f(x) \text{ terhadap } H_0 \quad \hat{f}_h(x) \neq f(x) \text{ untuk suatu nilai } h$$

Kemudian digunakan uji rasio kemungkinan (likelihood ratio test)

$f(x)/\hat{f}_h(x)$.

Untuk nilai h yang baik(optimal), statistik $f(x)/\hat{f}_h(x)$ mendekati satu.

Jadi $E[\log (f(x)/\hat{f}_h)(x)]$ mendekati nol. Persamaan

$$E \left| \log \left(\frac{f(x)}{\hat{f}_h} \right) (x) \right| = \int \log (f/\hat{f}_h)(x) f(x) dx$$

dikenal sebagai Kulback-Liebler Information yang memenuhi sifat suatu jarak sehingga dinotasikan dengan $d_{KL}(f, \hat{f}_h)$. Karena fungsi densitas $f(x)$ tidak diketahui, maka nilai $d_{KL}(f, \hat{f}_h)$ tidak dapat dihitung dari data. Secara intuitif dapat bahwa jika $d_{KL}(f, \hat{f}_h)$ mendekati nol, maka H_0 dapat diterima.

Untuk itu dicari nilai h yang meminimumkan $d_{KL}(f, \hat{f}_h)$

Dari beberapa teknik mencari nilai h optimal di atas, nilai h terbaik untuk menaksir densitas kernel suatu observasi adalah minimum antara nilai-nilai h optimal (hopt) dan nilai h oversmoothing (hos).

3.3 Taksiran Parameter dan Uji Statistik

Setelah menaksir fungsi densitas kernel dengan nilai h -nya dan fungsi kernel gaussian. Langkah selanjutnya yaitu menaksir nilai parameternya. Hal ini disebabkan adanya perbedaan parameter pada fungsi-fungsi densitas yang coba untuk ditaksir atau dilakukan pendekatan. Dalam bagian ini akan

dibahas penaksiran titik suatu parameter fungsi densitas dengan metode *moment* dan metode *maximum likelihood*.

3.3.1 Taksiran Paramater

a. Metode Moment

Misal sampel random $X_1, X_2, X_3, \dots, X_n$ dari fungsi densitas $f(x; \theta), x \in R^+$, dimana θ adalah vektor dari parameter p yang tidak diketahui. Pandang suatu moment teoritik $\mu^r = \int_0^\infty x^r f(x; \theta) dx$; $r = 1, 2, 3, \dots$, adalah fungsi dari θ . Pasangkan moment teoritik tersebut dengan moment sampel $\hat{\mu}^r = \frac{1}{n} \sum_{i=1}^n X_i^r$; $r = 1, 2, \dots$ maka akan didapat sistem persamaan untuk mencari nilai θ .

Contoh 1: Pareto(α, θ)

$$f(x; \alpha, \theta) = \frac{\alpha}{\theta} \left(\frac{\theta}{x + \theta}\right)^{\alpha+1}; x > 0$$

$$\hat{\mu}_r = E(X^r) = \frac{\theta^r r}{r-1(\alpha-1)}; r = 1, 2, \dots < \alpha$$

Pasangkan moment teoritis di atas dengan moment sampel, maka akan didapat:

$$\frac{\tilde{\theta}}{\tilde{\alpha} - 1} = \hat{\mu}_1 = \bar{X} = \tilde{\theta} = \bar{X}(\tilde{\alpha} - 1)$$

dan jika nilai $\alpha > 2$

$$\frac{2\tilde{\theta}^2(\tilde{\alpha} - 1)}{(\tilde{\alpha} - 2)} = \hat{\mu}'_2 = \frac{2\bar{X}^2(\tilde{\alpha} - 1)}{(\tilde{\alpha} - 2)} = \hat{\mu}'_2$$

Maka nilai:

$$\tilde{\alpha} = \frac{2(\hat{\mu}'_2 - \bar{X}^2)}{\hat{\mu}'_2 - 2\bar{X}^2} \text{ dan } \tilde{\theta} = \frac{\bar{X}\hat{\mu}_2}{(\hat{\mu}_2 - 2\bar{X}^2)}$$

b. Metode Maximum Likelihood

Misalkan sampel acak $X_1, X_2, X_3, \dots, X_n$ dari distribusi dengan fungsi distribusi kumulatif (cdf): $F(x; \theta), x \in R^+$ dan variabel θ adalah suatu vektor di R^p .

Misal fungsi likelihood $L(\theta; x_1, x_2, x_3, \dots, x_n) = L(\theta)$, maka untuk

- $F(\cdot; \theta)$ diskrit, fungsi likelihoodnya

$$L(\theta) = \prod_{i=1}^n [F(x_i; \theta) - F(x_{i-1}; \theta)]$$

Dengan kata lain,

$$L(\theta) = \prod_{i=1}^n [f(x_i; \theta)]$$

- $F(\cdot; \theta)$ kontinu, Misalkan a_i suatu elemen di dalam interval pendek $(a_1, b_1]$, maka

$$L(\theta) = \prod_{i=1}^n [F(b_i; \theta) - F(a_i; \theta)]$$

Jika nilai $[\Delta_1 = b_1 - a_1]$ kecil maka $F(b_i; \underline{\theta}) - F(a_i; \underline{\theta}) \approx f(x_i; \underline{\theta})\Delta_i$ dimana $f(\cdot; \underline{\theta})$ adalah fungsi densitas. Selanjutnya didapat:

$$L(\underline{\theta}) = \prod_{i=1}^n f(x_i; \underline{\theta})\Delta_i \approx \prod_{i=1}^n f(x_i; \underline{\theta})$$

Dengan memaksimalkan $L(\underline{\theta})$ atau $\ell(\underline{\theta}) = \ln L(\underline{\theta})$ akan kita peroleh nilai taksiran $\underline{\theta}$ yaitu nilai $\hat{\underline{\theta}}$.

Didefinisikan fungsi nilai dari $\underline{\theta}$ sebagai berikut:

$$S_j(\underline{\theta}) = \frac{\partial}{\partial \theta_j} \ell(\underline{\theta}) = \frac{\partial}{\partial \theta_j} \ln L(\underline{\theta}) = \frac{\frac{\partial}{\partial \theta_j} L(\underline{\theta})}{L(\underline{\theta})} \quad ; \quad j = 1, 2, \dots, p$$

Maksimalisasi diperoleh dengan mencari solusi dari sistem persamaan:

$$S_j(\underline{\theta}) = 0 \quad ; \quad j = 1, 2, \dots, p$$

Untuk beberapa kasus dalam menaksir parameter fungsi densitas diperlukan metode aproksimasi *Newton-Raphson* untuk mencari solusi sistem persamaan di atas.

Berikut adalah contoh penggunaan metode *Newton-Raphson* untuk menaksir parameter fungsi densitas.

Taksiran parameter untuk Distribusi Lognormal (μ, σ^2) :

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2 x}} \exp\left(-\frac{\ln x - \mu^2}{2\sigma^2}\right) ; x > 0$$

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \prod_{i=1}^n x_i \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2\right]$$

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{n}{2} - \sum_{i=1}^n \ln x_i - \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2$$

Maka, dengan mengambil nilai $S_1=S_2=0$ didapat:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln x_i \quad \text{dan} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\ln x_i - \hat{\mu})^2$$

Dalam penulisan ini akan dibahas penaksiran titik suatu parameter fungsi densitas metode maximum likelihood, yang nilai parameternya dapat dicari langsung dengan program FIT.

3.3.2 Uji Statistik

Setelah kita mendapatkan plot fungsi densitas Kernel dan nilai parameternya, langkah selanjutnya mencari kandidat kelas-kelas fungsi densitas dengan membandingkan taksiran densitas kernel dengan beberapa kelas fungsi densitas yang serupa. Dari beberapa kandidat yang diambil, langkah berikutnya yaitu melakukan uji statistik taksiran densitas kernel terhadap kandidat-kandidat fungsi densitas.

Pengujian kandidat fungsi densitas dengan melakukan salah satu uji "Goodness of Fit".

Penulisan ini menggunakan uji rasio likelihood dalam menguji taksiran densitas terhadap kandidat-kandidat fungsi densitas.

Berikut adalah beberapa metode untuk melakukan pengujian statistik, diantaranya:

- Uji Ratio Likelihood

Uji ini dilakukan manakala ditemukan ada kandidat fungsi densitas yang lebih sederhana (banyaknya parameter lebih sedikit) dibandingkan dengan kandidat fungsi densitas yang lainnya. Jika hal tersebut terjadi maka Uji Ratio Likelihood akan memberikan koreksi dengan *penalty*.

Adapun hipotesisnya sebagai berikut:

H_0 : fungsi distribusi yang lebih sederhana lebih baik (A)

H_1 : fungsi distribusi yang lebih kompleks lebih baik (B)

Statistik uji-nya adalah:

$$X^2 = 2|\text{NLL}_A - \text{NLL}_B| \sim \chi^2_{\Delta r}, \quad \text{dimana}$$

NLL : *Negatif Likelihood* masing-masing fungsi densitas yaitu negatif dari besarnya fungsi likelihood pada titik maksimum.

r : beda banyaknya parameter pada A dan B

Kriteria penolakan: H_0 ditolak jika $\chi^2 > \text{nilai tabel}$.

- Schwartz Bayesian Criterion (SBC)

Yaitu kriteria pemberian penalty yang ditambahkan pada NLL Uji Likelihood Ratio untuk membandingkan dua kandidat fungsi distribusi yang berbeda jumlah parameternya.

Besarnya *penalty* (p) yang diberikan adalah $p = r \log(N/2\pi)$, dimana :

r : banyaknya parameter

N : ukuran sampel

Kandidat fungsi distribusi yang dipilih sebagai taksiran fungsi distribusi observasi adalah yang mempunyai *score*: $NLL + \text{penalty}$ lebih kecil.

BAB IV

APLIKASI METODE KERNEL UNTUK PENAKSIRAN POLA PENGUNAAN LALU-LINTAS INTERNET

Dalam bab ini akan dibahas mengenai tahapan yang diperlukan untuk mengetahui pola penggunaan lalu-lintas internet dengan menggunakan penaksir densitas kernel, yaitu: pengambilan data, pengolahan data, dan analisa data.

4.1 Data Penggunaan Internet

Data penggunaan internet adalah berupa paket (package) yang melalui jaringan router, yang ditransmisikan di router jaringan. Paket diambil atau ditangkap menggunakan perangkat lunak atau software TCP dump dan diterjemahkan kembali menggunakan software *Ethereal*. Paket yang diambil kemudian dipilih beberapa informasi terkait di dalamnya antara lain waktu paket tersebut ditangkap, tujuan paket (destination), asal paket (source), jenis protokol yang digunakan, panjang paket, waktu kedatangan paket.

Untuk itu akan dijelaskan tahapan bagaimana data diambil, dipilih, diproses dan dianalisa sehingga memiliki informasi bagi penelitian tugas akhir ini. Berikut adalah tahapan yang dimaksud.

4.1.2 Pengambilan Data

Pengambilan atau penangkapan data (data capture) melalui dua tahapan, yaitu:

1. Penangkapan paket (package capture).

Paket yang melewati titik gerbang jaringan router, ditangkap menggunakan software *tcp-dump*.

Tcp-dump menangkap setiap paket yang melewati router dan mengumpulkannya dalam memori komputer. Paket yang melewati router ini masih dalam format biner. Karena keterbatasan memori komputer dan kemungkinan perhitungan selanjutnya, paket yang ditangkap dan dikumpulkan, diatur tiap lima belas (15) menit.

Pemilihan besarnya paket untuk tiap frame berdasarkan tujuan dan asal paket tersebut (keluar dan ke dalam router (berlakusebaliknya)). Misal alamat asal paket di 192.168.215 ke alamat tujuan paket 66.23.171 dengan ukuran besar paket yakni 105 byte.

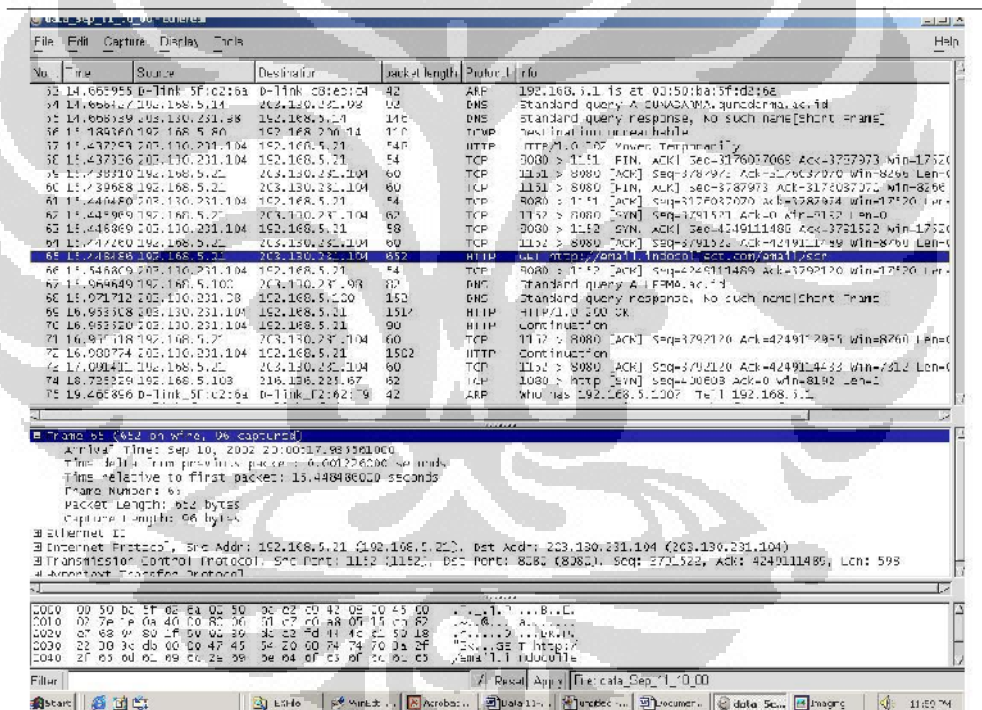
Penjumlahan data terpilih tiap frame per 15 menit pengambilan data. Misal total paket pada pukul 10:15 WIB di tanggal 11 oktober 2002 dengan ukuran paket sebesar 12288 byte.

2. Penerjemahan paket.

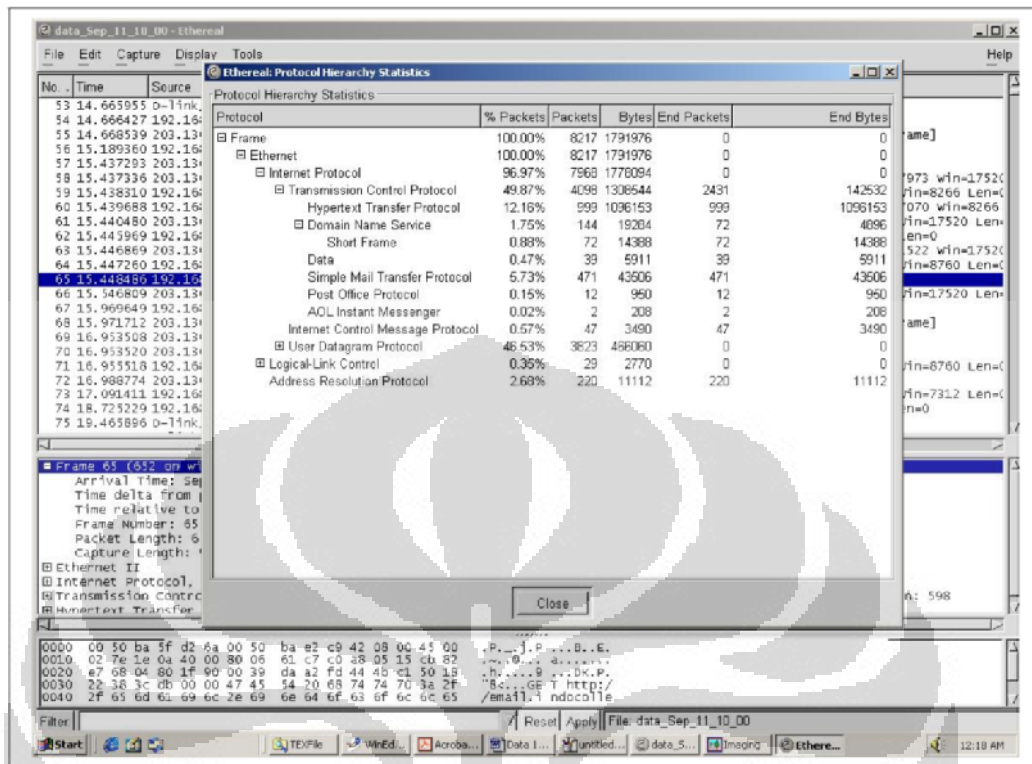
Paket yang ditangkap dan dikumpulkan oleh *tcp-dump* berisikan informasi mengenai data yang ditransmisikan di Internet, diterjemahkan oleh

software *Ethereal*. Informasi paket yang ditampilkan oleh *Ethereal*, antara lain waktu paket tersebut ditangkap, tujuan paket (destination), asal paket (source), jenis protokol yang digunakan, panjang paket, waktu kedatangan paket, dan informasi tambahan lainnya.

Sebagai gambaran mengenai bagaimana suatu data dalam bentuk paket-paket diambil dan diterjemahkan serta disummarykan menggunakan perangkat lunak *Ethereal*, diilustrasikan pada gambar 4.1 dan gambar 4.2.



Gambar 4.1. Tampilan informasi paket dalam Ethereal



Gambar 4.2 Tampilan protocol summary dalam Ethereal

Keterangan gambar 4.2: Dengan menggunakan protokol summary pada Ethereal, dapat dihitung total jumlah paket untuk empat jenis protokol (HTTP, SMTP, TCP, ICMP) yang digunakan yang ditangkap atau diambil pada pukul 08.00 -17.00.

4.2 Pengolahan Data

Untuk mengolah data yang didapat, penulisan tugas akhir ini menggunakan perangkat lunak pengolahan data yang disebut R.

Ada tiga tahapan yang diperlukan untuk pengolahan data, yaitu:

- Plot taksiran densitas. Dalam tahapan ini, Total jumlah paket dari empat jenis protokol yang digunakan, kemudian dicari parameter

pemulusan h terbaik dan kemudian dibuat plot taksiran densitasnya. Teknik pemilihan parameter h dibahas pada bab III dan membuat plot densitasnya, menggunakan sintak tertentu dalam **R**. Adapun sintaks yang digunakan adalah:

```
plot (density(X, bw = nilai h), xlab = "paket (Mega byte)", main="keterangan plot")
```

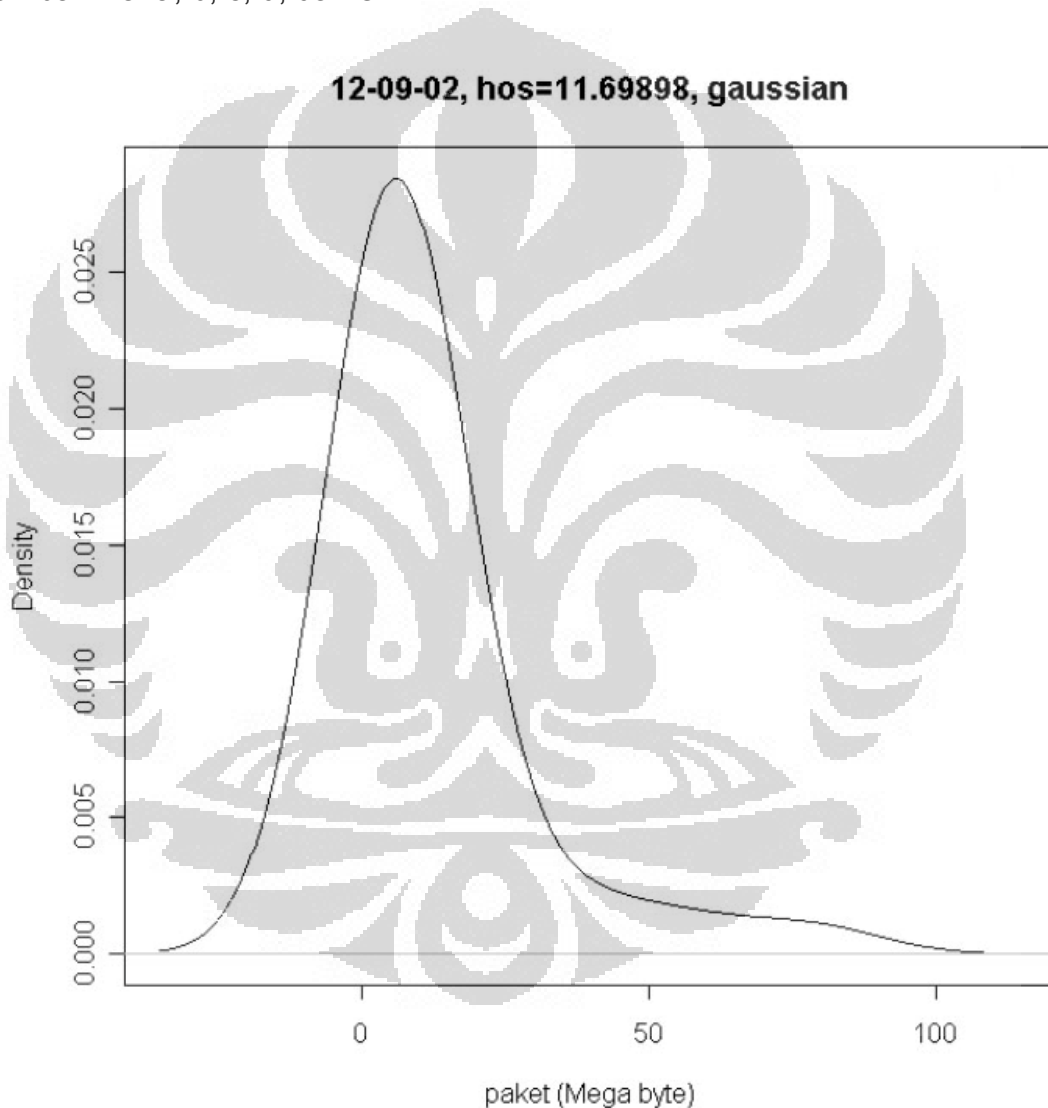
- Penaksiran parameter. Dalam bagian ini Apabila kita mendapatkan kandidat fungsi densitas untuk taksiran densitas data, penaksiran nilai parameter dikerjakan dengan menggunakan perangkat lunak lainnya, yaitu FIT.
- Uji statistik. aksiran fungsi densitas yang mempunyai kandidat fungsi densitas dan nilai parameternya, kemudian diuji dengan menggunakan uji statistik rasio likelihood, untuk mengetahui taksiran densitas tersebut, mendekati suatu kelas fungsi densitas tertentu. Dibahas pada bab III.

4.3 Analisa Data

Pada subbagian ini beberapa tahapan yang perlu diperhatikan untuk dapat menganalisa data dengan baik, seperti berikut ini.

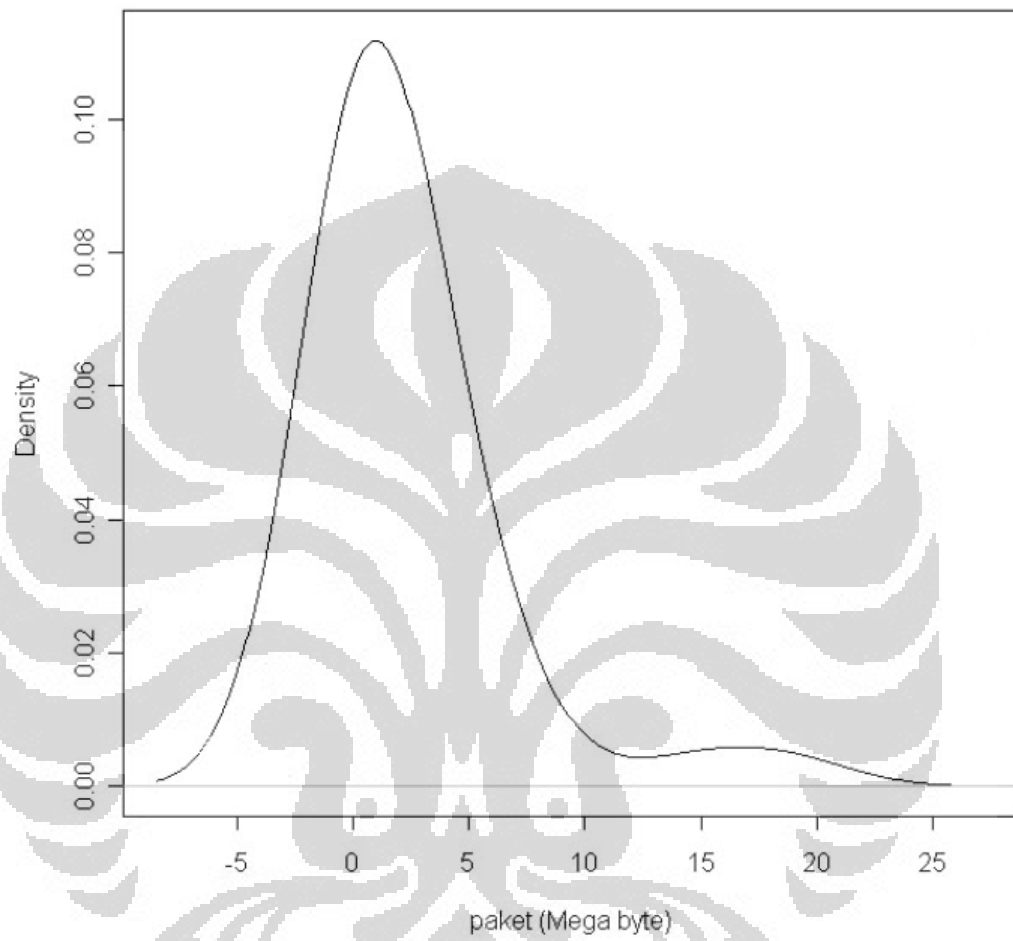
4.3.1 Kandidat fungsi densitas dan nilai parameter

Dengan parameter pemulusan h terbaik dan fungsi kernel gaussian diperoleh plot taksiran densitas untuk data tanggal 12, 14, 17, 24 dan 27 September 2002, dengan total jumlah paket yang lebih besar 0. Seperti pada gambar 4.3. a, b, c, d, dan e.



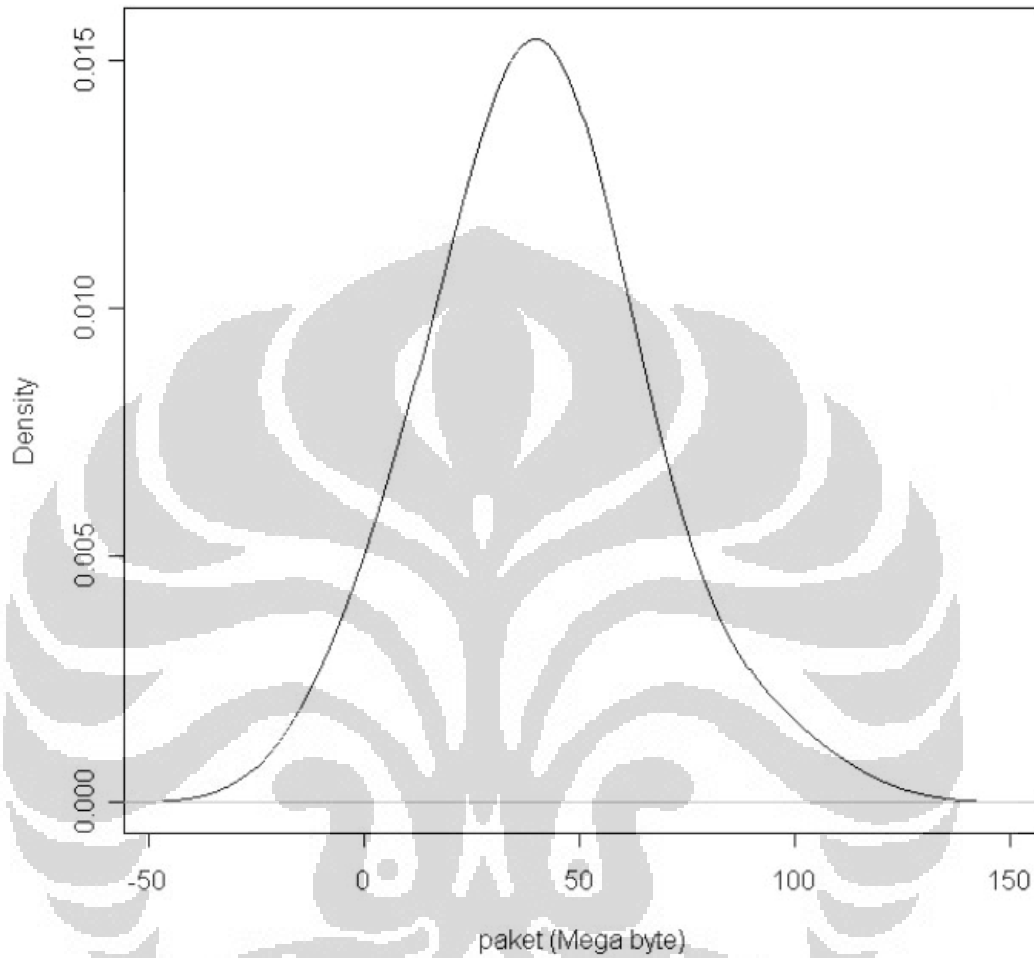
Gambar 4.3.a Taksiran Densitas Penggunaan Lalulintas Internet pada
Tanggal 12 September 2002

14-09-02, hos= 2.820781, gaussian



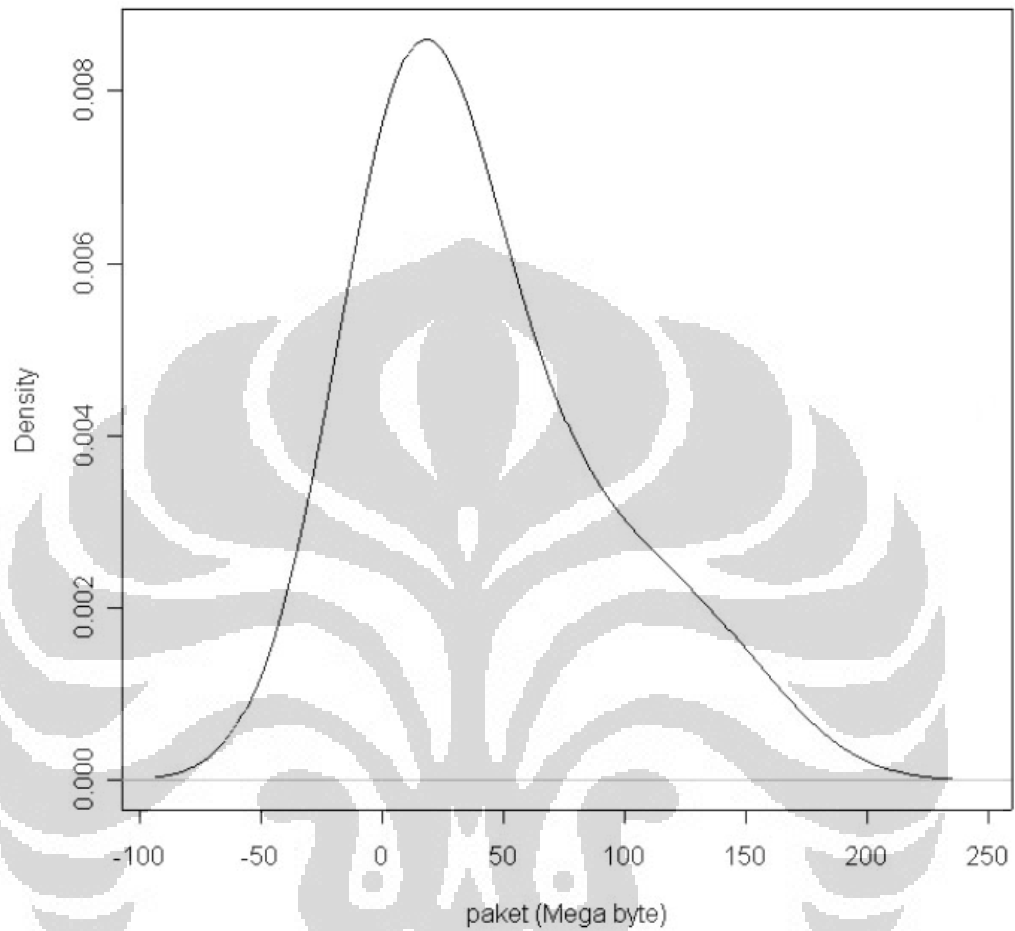
Gambar 4.3.b Taksiran Densitas Penggunaan Lalulintas Internet pada Tanggal 14 September 2002

17-09-02, hos= 15.88315, gaussian



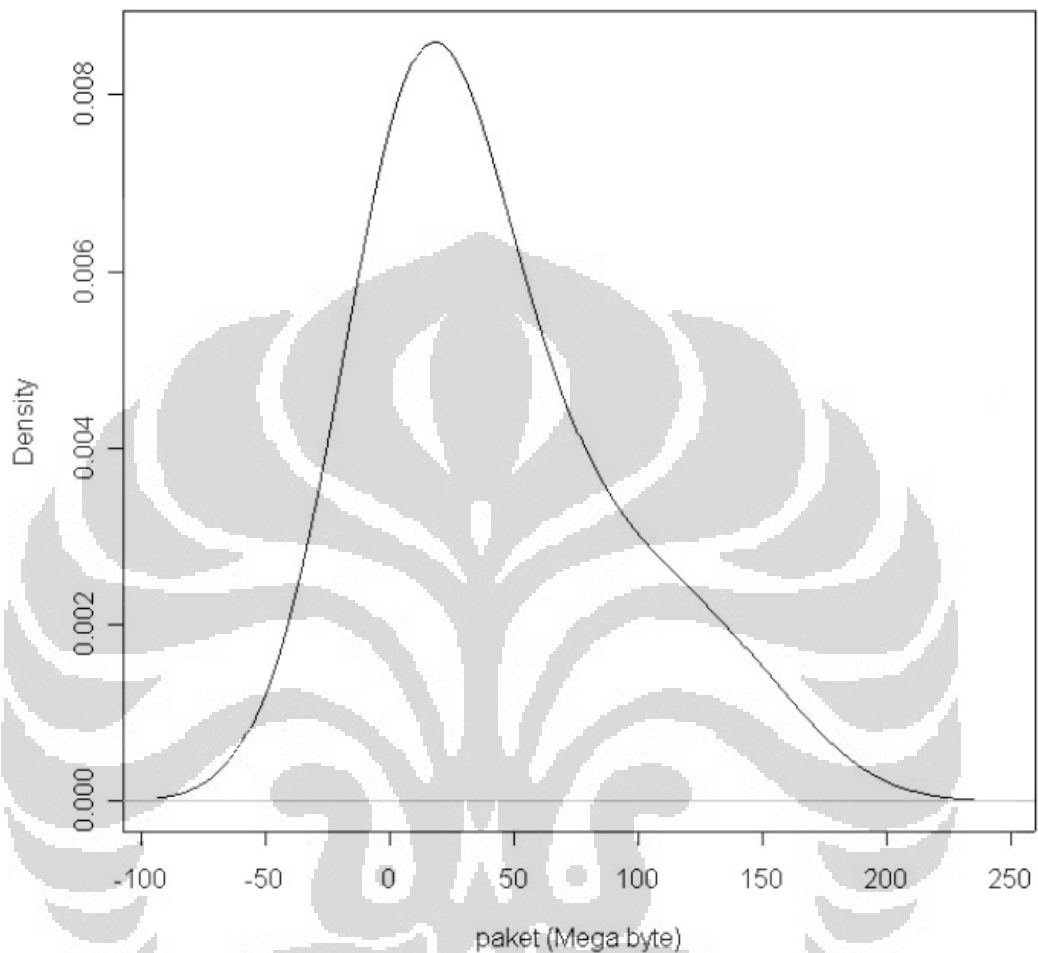
Gambar 4.3.c Taksiran Densitas Penggunaan Lalulintas Internet pada
Tanggal 17 September 2002

24-09-02, hos=31.18851, gaussian



Gambar 4.3.d Taksiran Densitas Penggunaan Lalulintas Internet pada
Tanggal 24 September 2002

27-09-02, hos= 31.18851, gaussian



Gambar 4.3.e Taksiran Densitas Penggunaan Lalulintas Internet pada Tanggal 27 September 2002

Pada gambar 4.3 a, b, c, d, dan e, masing-masing tanggal memiliki kandidat fungsi densitasnya, yaitu:

1. Data penggunaan tanggal 12 Sept 2002, dengan kandidat fungsinya yaitu:

- Lognormal, dengan nilai parameter

$$\hat{\mu} = 0.447134; \hat{\sigma} = 1.49963; NLL = 9.13685$$

- Generalized Pareto, dengan nilai parameter

$$\hat{\alpha} = 2.26748; \hat{\theta} = 6.10194; \hat{\tau} = 1.000; NLL = 9.6257$$

- Eksponensial, dengan nilai parameter

$$\hat{\theta} = 4.81422; NLL = 18.1173$$

2. Data penggunaan tanggal 14 Sept 2002, dengan kandidat fungsinya yaitu:

- Lognormal, dengan nilai parameter

$$\hat{\mu} = -1.28543; \hat{\sigma} = 1.58106; NLL = 7.87556$$

- Gamma, dengan nilai parameter

$$\hat{\alpha} = 0.0894484; \hat{\theta} = 10.7892; \hat{\tau} = ; NLL = 6.99925$$

- Eksponensial, dengan nilai parameter

$$\hat{\theta} = 9.65073; NLL = 19.4728$$

3. Data penggunaan tanggal 17 Sept 2002, dengan kandidat fungsinya yaitu:

- Lognormal, dengan nilai parameter

$$\hat{\mu} = 2.01268; \hat{\sigma} = 1.16223; NLL = 8.21121$$

- Gamma, dengan nilai parameter

$$\hat{\alpha} = 0.34959; \hat{\theta} = 42.0581; NLL = 7.66683$$

- Eksponensial, dengan nilai parameter

$$\hat{\theta} = 14.7033; NLL = 19.60401$$

4. Data penggunaan tanggal 24 Sept 2002, dengan kandidat fungsinya yaitu:

- Lognormal, dengan nilai parameter

$$\hat{\mu} = -0.120633; \hat{\sigma} = 1.32923; NLL = 7.07574$$

- Gamma, dengan nilai parameter

$$\hat{\alpha} = -0.26083; \hat{\theta} = 10.4048; NLL = 6.40305$$

- Eksponensial, dengan nilai parameter

$$\hat{\theta} = 2.14426; NLL = 10.5306$$

5. Data penggunaan tanggal 27 Sept 2002, dengan kandidat fungsinya yaitu:

- Lognormal, dengan nilai parameter

$$\hat{\mu} = 1.81138; \hat{\sigma} = 1.34904; NLL = 9.08410$$

- Gamma, dengan nilai parameter

$$\hat{\alpha} = 0.193378; \hat{\theta} = 78.6055; NLL = 8.39621$$

- Eksponensial, dengan nilai parameter

$$\hat{\theta} = 15.2005; NLL = 12.7446$$

4.3.2 Uji Statistik dan Penarikan Kesimpulan

Setelah mendapatkan kandidat fungsi densitasnya dan nilai taksiran parameternya, selanjutnya dilakukan uji statistik untuk masing masing data per tanggal penggunaan internet.

1. Untuk data penggunaan pada tanggal 12 Sept 2002, distribusi eksponensial mempunyai satu parameter, sedangkan distribusi Generalized Pareto, dan Lognormal mempunyai dua parameter. Antara dua distribusi *Generalized Pareto* dan *Lognormal*, yang dipilih adalah distribusi Lognormal karena memiliki nilai *Negatif Likelihood (NLL)* yang paling kecil dibandingkan *Generalized Pareto*, yang disajikan pada tabel 4.1.

Tabel 4.1 Score *NLL* Data Tanggal 12 September 2002

Distribusi	<i>NLL</i>	Penalty	Score (<i>NLL</i> + penalty)
Eksponensial	18.1173	0.77	18.8873
Log-normal	9.13685	1.54	10.67685

Kemudian akan dibandingkan fungsi distribusi *Generalized Pareto* dengan *Eksponensial*, karena banyaknya parameter berbeda maka akan dilakukan uji rasio likelihood dan SBC, dimana:

H_0 : Distribusi Eksponensial lebih baik (A)

H_1 : Distribusi Lognormal lebih baik (B)

Statistik uji dimana, $\chi^2 = 2(NLLA - NLLB) = 2(18.1173 - 9.13685) = 17.9609$ dengan kriteria penolakan $\chi^2 = 17.9609 > \text{nilai tabel}(\chi^2_{r=1}) = 3.84$ maka H_0 ditolak. Dengan kata lain bahwa distribusi Lognormal lebih baik.

Karena banyaknya parameter tidak sama, maka masing-masing distribusi akan dilihat mana yang lebih baik dari score-nya dengan menambahkan *penalty* terhadap *NLL*. Dimana nilai $penalty = 0.77r$ didapat.

Karena $score\ distribusi\ Lognormal < score\ distribusi\ Eksponensial$, jadi distribusi Lognormal lebih baik.

Maka dapat disimpulkan bahwa **data penggunaan internet pada tanggal 12 cenderung berdistribusi Lognormal.**

2. Untuk data penggunaan pada tanggal 14 Sept 2002, distribusi eksponensial mempunyai satu parameter, sedangkan distribusi Gamma, dan Lognormal mempunyai dua parameter. Antara distribusi Gamma dan Lognormal, yang dipilih adalah distribusi Gamma karena mempunyai *Negatif Likelihood (NLL)* yang paling kecil, yang disajikan pada tabel 4.2.

Tabel 4.2 Score *NLL* Data Tanggal 14 September 2002

Distribusi	<i>NLL</i>	Penalty	Score (<i>NLL</i> + penalty)
Eksponensial	19.4728	0.77	20.2428
Gamma	6.99925	1.54	8.5395

Selanjutnya akan dibandingkan fungsi distribusi Gamma dengan eksponensial, karena banyaknya parameter berbeda maka akan dilakukan uji likelihood ratio dan SBC, dimana:

H_0 : Distribusi Eksponensial lebih baik (A)

H_1 : Distribusi Gamma lebih baik (B)

Statistik uji dimana, $\chi^2 = 2(NLLA - NLLB) = 2(19.4728 - 10.7892) = 4.3418$ dengan kriteria penolakan $\chi^2 = 4.3418 > \text{nilai tabel}(\chi^2_{r=1}) = 3.84$ maka H_0 ditolak. Dengan kata lain bahwa distribusi Gamma lebih baik.

Karena banyaknya parameter tidak sama, maka masing-masing distribusi akan dilihat mana yang lebih baik dari score-nya dengan menambahkan *penalty* terhadap *NLL*. Dimana nilai *penalty* yang ditambahkan adalah $\text{penalty} = 0.77r$. Karena $\text{score distribusi Gamma} < \text{score distribusi Eksponensial}$, maka distribusi Gamma lebih baik.

Oleh karena itu dapat disimpulkan bahwa **data penggunaan internet pada tanggal 14 Sept 2002 cenderung berdistribusi Gamma.**

3. Untuk data penggunaan tanggal 17 Sept 2002, distribusi eksponensial mempunyai satu parameter, sedangkan distribusi Gamma, dan Lognormal mempunyai dua parameter. Antara distribusi Gamma dan Lognormal, dipilih distribusi Gamma karena mempunyai Negatif Likelihood (NLL) yang paling kecil, yang disajikan pada tabel 4.3.

Tabel 4.3 Score *NLL* Data Tanggal 17 September 2002

Distribusi	<i>NLL</i>	Penalty	Score (<i>NLL</i> + penalty)
Eksponensial	9.60401	0.77	10.37401
Gamma	7.66683	1.54	9.20683

Selanjutnya akan dibandingkan fungsi distribusi Gamma dengan eksponensial, karena banyaknya parameter berbeda maka akan dilakukan uji likelihood ratio dan SBC, dimana:

H_0 : *Distribusi Eksponensial lebih baik (A)*

H_1 : *Distribusi Gamma lebih baik (B)*

Statistik uji dimana, $\chi^2 = 2(NLLA - NLLB) = 2(9.60401 - 8.21121) = 2.78568$ dengan kriteria penolakan $\chi^2 = 2.78568 > \text{nilai tabel}(\chi^2_{r=1}) = 3.84$ maka H_0 ditolak. Dengan kata lain bahwa distribusi Eksponensial lebih baik.

Karena banyaknya parameter tidak sama, maka masing-masing distribusi akan dilihat mana yang lebih baik dari score-nya dengan menambahkan *penalty* terhadap *NLL*. Dimana nilai *penalty* yang ditambahkan adalah $\text{penalty} = 0.77r$. Karena *score distribusi Gamma* < *score distribusi Eksponensial*, maka distribusi Gamma lebih baik. Oleh karena itu dapat disimpulkan bahwa **data penggunaan internet pada tanggal 17 Sept 2002 cenderung berdistribusi Gamma.**

4. Untuk data penggunaan pada tanggal 24 Sept 2002, distribusi eksponensial mempunyai satu parameter, sedangkan distribusi Gamma, dan Lognormal mempunyai dua parameter. Antara distribusi Gamma dan Lognormal, dipilih distribusi Gamma karena mempunyai Negatif Likelihood (NLL) yang paling kecil, yang disajikan pada tabel 4.4.

Tabel 4.4 Score *NLL* Data Tanggal 24 September 2002

Distribusi	<i>NLL</i>	Penalty	Score (<i>NLL</i> + penalty)
Eksponensial	10.5306	0.77	11.3006
Gamma	6.40305	1.54	7.94305

Selanjutnya akan dibandingkan fungsi distribusi Gamma dengan eksponensial, karena banyaknya parameter berbeda maka akan dilakukan uji likelihood ratio dan SBC, dimana:

H_0 : Distribusi Eksponensial lebih baik (A)

H_1 : Distribusi Gamma lebih baik (B)

Statistik uji dimana, $\chi^2 = 2(NLLA - NLLB) = 2(10.5306 - 7.07574) = 6.90972$ dengan kriteria penolakan $\chi^2 = 2.78568 > \text{nilai tabel}(\chi^2_{r=1}) = 3.84$ maka H_0 ditolak. Dengan kata lain bahwa distribusi Gamma lebih baik.

Karena banyaknya parameter tidak sama, maka masing-masing distribusi akan dilihat mana yang lebih baik dari score-nya dengan

menambahkan *penalty* terhadap *NLL*. Dimana nilai *penalty* yang ditambahkan adalah $penalty = 0.77r$. Karena *score distribusi Gamma* < *score distribusi Eksponensial*, maka distribusi Gamma lebih baik.

Oleh karena itu dapat disimpulkan bahwa **data penggunaan internet pada tanggal 24 Sept 2002 cenderung berdistribusi Gamma.**

5. Untuk data penggunaan pada tanggal 27 Sept 2002, distribusi eksponensial mempunyai satu parameter, sedangkan distribusi Gamma, dan Lognormal mempunyai dua parameter. Antara distribusi Gamma dan Lognormal, dipilih distribusi Gamma karena mempunyai Negatif Likelihood (NLL) yang paling kecil, yang disajikan pada tabel 4.5.

Tabel 4.5 Score *NLL* Data Tanggal 27 September 2002

Distribusi	<i>NLL</i>	Penalty	Score (<i>NLL</i> + penalty)
Eksponensial	12.7446	0.77	13.5146
Gamma	8.39621	1.54	9.93624

Selanjutnya akan dibandingkan fungsi distribusi Gamma dengan eksponensial, karena banyaknya parameter berbeda maka akan dilakukan uji likelihood ratio dan SBC, dimana:

H_0 : Distribusi Eksponensial lebih baik (A)

H_1 : Distribusi Gamma lebih baik (B)

Statistik uji dimana, $\chi^2 = 2(NLLA - NLLB) = 2(12.7446 - 8.39621) = 8.6978$
dengan kriteria penolakan $\chi^2 = 8.6978 > \text{nilai tabel}(\chi^2_{r=1}) = 3.84$ maka H_0
ditolak. Dengan kata lain bahwa distribusi Gamma lebih baik.

Karena banyaknya parameter tidak sama, maka masing-masing distribusi akan dilihat mana yang lebih baik dari score-nya dengan menambahkan *penalty* terhadap NLL. Dimana nilai *penalty* yang ditambahkan adalah $\text{penalty} = 0.77r$. Karena $\text{score distribusi Gamma} < \text{score distribusi Eksponensial}$, maka distribusi Gamma lebih baik.

Oleh karena itu dapat disimpulkan bahwa **data penggunaan internet pada tanggal 27 Sept 2002 cenderung berdistribusi Gamma.**

BAB V

PENUTUP

5.1 Kesimpulan

Dari pembahasan pada bab-bab sebelumnya dapat diambil kesimpulan bahwa:

- data penggunaan internet pada tanggal 12 cenderung berdistribusi Lognormal.
- data penggunaan internet pada tanggal 14 Sept 2002 cenderung berdistribusi Gamma.
- data penggunaan internet pada tanggal 17 Sept 2002 cenderung berdistribusi Gamma.
- data penggunaan internet pada tanggal 24 Sept 2002 cenderung berdistribusi Gamma.
- data penggunaan internet pada tanggal 27 Sept 2002 cenderung berdistribusi Gamma.

Dengan melihat pada kesimpulan pengujian di atas bahwa secara keseluruhan penggunaan internet pada bulan September 2002 tidak adanya pola seragam walaupun kebanyakan berdistribusi Gamma.

5.2 Saran

Dalam rangka meningkatkan kualitas dan ruang lingkup permasalahan, saran yang sebaiknya dapat dilakukan ke depan:

- Sampel pengambilan yang lebih besar, untuk tiap tanggal pada tiap bulannya. Karena tanggal penggunaan lalu-lintas Internet yang ditaksir hanya 5 hari, dan nilai observasi yang sedikit (37 observasi). Disarankan penaksiran fungsi densitas penggunaan Internet dalam kurun waktu yang lebih lama (data untuk tiap hari selama 1 bulan penuh). Dan selisih titik waktu pengambilan data menjadi lebih sempit (kurang dari 15 menit) sehingga didapat titik waktu pengambilan data yang lebih banyak. Namun dalam praktiknya, langkah di atas membutuhkan biaya yang besar antara lain waktu yang dibutuhkan untuk mengumpulkan data menjadi lebih lama, kapasitas penyimpanan data yang sangat besar, dan pengolahan data yang lebih lama.
- Proses filterisasi data yang lebih baik. Tidak dapat dipungkiri bahwa paket yang melewati internet di filter menggunakan perangkat lunak tertentu dan tidak menutup kemungkinan proses filterisasi tidak optimal. Untuk itu perlu adanya program khusus dalam tahapan filterisasi ini.

DAFTAR PUSTAKA

1. Wolfgang Hardle. *Smoothing Technique: with Implementation in S*. Springer-Verlag, 1990.
2. Behrouz A. Forouzan. *TCP/IP Protocol Suite*. McGraw Hill, 2000.
3. Yuniah Wati. *Penaksir Kernel Densitas Non-parameterik dengan Metode Kernel*. Jurusan Matematika FMIPA UI, 2001.
4. Rianti S, Netty S. *Workshop Loss Distribution*. Jurusan Matematika FMIPA UI, 2001.
5. Kevin Lai, Mary Baker. *Nettimer: A Tool for Measuring Bottleneck Link Bandwidth*. Departemen of Computer Science, Stanford University, April 2001.
6. www.gtreach.com/globalstats. *Global Internet Statistics by Language*, September 2001.

Lampiran 1

Metode *Sheater & Jones (1991)* untuk pemilihan bandwidth h untuk penaksir kernel di software R:

```
function (x, nb=1000, lower=0.1*hmax, upper=hmax, method=c("ste:", "dpi"))
{
  fSD<- function(h, x, alph2, c1, n, d) (c1/SDh(x, alph2*h^(5/7). N, d))^(1/5)-h
  SDh<-function(x, h, n, d) .C("band_phi4_bin", as.integer(n),
  as.integer(length(x)), as.double(d), x, as.double(h),
  u=double(1), PACKAGE="base")$u
  method<-match.arg(method)
  if (!is.numeric(x) || !length(x))
  stop("invalid x")
  n<-length(x)
  storage.mode(x) <- "double"
  n<-length(x)
  Z <- .C("band_den_bin", as.integer(n), as.integer(nb), d=double(1), x,
  cnt=integer(nb), PACKAGE="base")
  D <- Z$d
  cnt <- as.integer(Z$cnt)
  hmax <- 1.144*sqrt(var(x))*n^(-1/5)
  scale <- min(sqrt(var(x)), IQR(x)/1.349)
  a <- 1.24* scale* n^(-1/7)
  b <- 1.23* scale* n^(-1/9)
  c1 <- 1/(2 * sqrt(pi) * n)
```

```

TD <- -TDh(cnt, b, n, d)
if (!is.finite(TD) || TD <=0)
stop("sample is to sparse to find TD")
alph2 <- 1.357 * (SDh(cnt, a, n, d)/TD)^(1/7)
If (method=="dpi")
res <- (c1/SDh(cnt, (2.394/(n * TD))^(1/7, n, d))^(1/5)
else {
if (!is.finite(alph2))
stop("sample is to sparse to find alph2")
if (fSD(lower, cnt, alph2, c1, n, d) * fSD(upper, cnt, alph2, c1, n, d) > 0)
stop("No solution in the specified range of bandwidths")
res <- uniroot(fSD, c(lower, upper), tol=0.1 * lower,
x=cnt, alph2=alph2, c1=c1, n=n, d=d)$root
}
res
}

```

Lampiran 2

Metode *Unbiased Cross Validation* untuk pemilihan bandwidth h untuk penaksir kernel di software R:

```
function (x, nb=1000, lower=0.1 * hmax, upper=hmax)
{
  funcv <- function(h, x, n, d) .C("band_ucv_bin", as.integer(n),
    as.integer(length(x), as.double(d), x, as.double(h),
    u = double(1), PACKAGE = "base")$u
  if (!is.numeric(x) || !length(x))
    stop("invalid x")
  n <- length(x)
  hmax <- 1.144 * sqrt(var(x)) * (-1/5)
  storage.mode(x) <- "double"
  Z <- .C("band_den_bin", as.integer(n), as.integer(nb), d = double(1),
    x, cnt = integer(nb), PACKAGE = "base")
  d <- Z$d
  cnt <- as.integer(z$cnt)
  h <- optimize(funcv, c(lower, upper), tol = 0.1 * lower, x = cnt, n = n,
    d = d)$minimum
  if (h < 1.1 * lower | h > upper - 0.1 * lower)
    warning("minimum occurred at one end of the range")
  h
}
```

Lampiran 3

Metode *Rule of Thumb* untuk pemilihan bandwidth h untuk penaksir kernel di software R:

```
function (x)
{
  hi <- sd(x)
  if (!lo <- min(hi, IQR(x)/1.34))
    lo <- hi || (lo <- abs(x[1])) || (lo <- 1)
  0.9 * lo * length(x) ^ (-0.2)
}
```

Lampiran 4

Algoritma penaksir densitas kernel untuk ukuran bandwidth h sama dan fungsi kernel yang berbeda:

```
bw <- bw.methode(data) ## sensible automatic choice
plot(density(data, bw = bw, n = 2^13),
     main = "same sd bandwidths, 7 different kernels")
for(i in 2: length(kernels))
  lines(density(precip, bw = bw, kern = kernels[i], n = 2^13), col = 1)
```


Lampiran 5

Algoritma penaksir densitas kernel untuk ukuran bandwidth h tertentu dan fungsi kernel

```
plot (density(data, bw = "sebutkan ukuran bandwidth), xlab = "", main = "judul
utama")
for(i in 2: length(kernels))
  lines(density(data, bw = bw, kern = kernels[i]), col = i)
legend(1.5, 0.4, legend = kernels, col = seq(kernels),
      lty =1, cex = 0.8, y.int = 1)
```