LAPORAN TUGAS BESAR

MATA KULIAH WI2002

LITERASI DATA DAN INTELEGENSI ARTIFISIAL

TAHUN 2025

Disusun oleh Kelompok 14 – Kelas 32

1. **Benedict Darrel Setiawan** **13524057**
2. **Marcel Luther Sitorus** **13524063**
3. **Muhammad Akmal** **13524099**
4. **Daniel Putra Rywandi S** **13524143**

MATA KULIAH WAJIB KURIKULUM

WAKIL REKTOR BIDANG AKADEMIK DAN KEMAHASISWAAN

INSTITUT TEKNOLOGI BANDUNG

BANDUNG

2025

DAFTAR ISI

BAB 1 – PENDAHULUAN

A. Latar Belakang

Di era digital saat ini, pertumbuhan layanan berbasis internet telah menjadikan server web sebagai infrastruktur penting bagi organisasi dan individu. Namun, ketergantungan yang meluas pada teknologi web ini juga membuat sistem rentan terhadap berbagai ancaman keamanan siber, termasuk serangan *Distributed Denial of Service* (DDoS), upaya *Brute-Force Login*, dan aktivitas jahat lainnya. Deteksi dan analisis yang efektif terhadap perilaku anomali dalam lalu lintas jaringan sangat penting untuk menjaga integritas sistem dan mencegah kerusakan.

Bidang analisis lalu lintas jaringan memanfaatkan data tingkat aliran dalam jumlah besar, yang merekam atribut seperti jumlah paket, durasi koneksi, kecepatan transfer data, dan penggunaan protokol. Dengan memeriksa atribut kuantitatif dan kategoris ini, menjadi mungkin untuk mengungkap pola tersembunyi yang membedakan perilaku normal dari serangan potensial.

Proyek ini bertujuan untuk menyelidiki hubungan antara fitur jaringan berbasis aliran dan keberadaan serangan siber. Dengan menganalisis lalu lintas yang ditangkap selama skenario jinak dan serangan, kami berupaya mengidentifikasi atribut mana yang berfungsi sebagai indikator kuat anomali. Secara khusus, fokusnya adalah pada pemodelan bagaimana karakteristik lalu lintas yang berbeda berkorelasi dengan perilaku jahat, menggunakan teknik pembelajaran mesin statistik dan sederhana.

B. Pertanyaan Penelitian

Berdasarkan latar belakang tersebut, kami merumuskan pertanyaan penelitian sebagai berikut:

1. Bagaimana metrik jaringan tingkat aliran (*flow-level network metrics*) seperti *packet counts*, ukuran byte, dan durasi aliran berkorelasi dengan keberadaan serangan siber dalam lalu lintas server web?
2. Dapatkah kita memprediksi anomali atau perilaku tidak teratur dalam lalu lintas jaringan berdasarkan fitur statistik berbasis aliran (*flow-based statistics*)?
3. Apakah terdapat pola tertentu pada fitur lalu lintas jaringan (misalnya durasi aliran, jumlah paket, ukuran rata-rata paket) yang dapat digunakan untuk mendeteksi keberadaan serangan DDoS?

C. Data

Untuk menjawab pertanyaan penelitian yang dirumuskan, proyek ini menggunakan *dataset* CIC-IDS2017 yang disediakan oleh Canadian Institute for Cybersecurity (CIC). Dataset CIC-IDS2017 berisi serangan umum yang jinak dan terkini, yang menyerupai data dunia nyata yang sebenarnya, sering disebut *analyzing packet captures* (PCAP). Dataset ini juga mencakup hasil analisis lalu lintas jaringan menggunakan CIC FlowMeter dengan aliran berlabel berdasarkan cap waktu, IP sumber dan tujuan, port sumber dan tujuan, protokol dan serangan.

Dataset CIC-IDS2017 dipilih karena representasinya yang komprehensif terhadap perilaku lalu lintas jaringan modern, yang menggabungkan aktivitas jinak dan berbagai skenario serangan siber dalam kondisi yang terkontrol dan realistis. Kumpulan data ini menyediakan metrik berbasis aliran yang terperinci, pelabelan yang ekstensif, dan komponen deret waktu, yang semuanya selaras erat dengan tujuan analitis proyek ini.

Meskipun CIC-IDS2017 secara resmi merupakan satu dataset terpadu, dataset ini terbagi menjadi koleksi harian yang menyimulasikan kondisi operasional. Pada hari-hari tertentu, dataset secara eksklusif terdiri dari lalu lintas normal yang bebas serangan, sementara pada hari-hari lainnya mencakup campuran aktivitas jinak dan berbahaya. Untuk memenuhi persyaratan penggunaan dua set data, kami memperlakukan lalu lintas normal yang direkam pada hari Senin, 3 Juli 2017, sebagai set data pertama (yang mewakili operasi server web dasar), dan lalu lintas serangan campuran yang direkam pada hari Rabu, 5 Juli 2017, sebagai set data kedua yang berbeda. Pemisahan ini dibenarkan secara struktural dan analitis, karena memungkinkan analisis komparatif antara perilaku jaringan yang umum dan lalu lintas yang terkena serangan.

URL Sumber Dataset: https://www.unb.ca/cic/datasets/ids-2017.html

Dataset 1 – Lalu-Lintas Normal

Spesifikasi data set pertama kami sebagai berikut:

- Sumber: Canadian Institute for Cybersecurity (CIC) – CIC-IDS2017 Dataset
- Nama File: Monday-WorkingHours.pcap_ISCX.csv
- Format: *Comma-Separated Values* (CSV)
- Ukuran: 262 MB
- Deskripsi: Simulasi rekaman aktivitas lalu-lintas jaringan normal tanpa aktivitas serangan.

Dataset 2 – Lalu-Lintas Serangan

Spesifikasi data set kedua kami sebagai berikut:

- Sumber: Canadian Institute for Cybersecurity (CIC) – CIC-IDS2017 Dataset
- Nama File: Wednesday-WorkingHours.pcap_ISCX.csv
- Format: *Comma-Separated Values* (CSV)
- Ukuran: 278 MB
- Deskripsi: Menggambarkan skenario serangan nyata yang dipadukan dengan perilaku normal.

D. Atribut Data

Kedua dataset menggunakan format *Comma-Separated Values* (CSV) yang memiliki *header* atribut yang sama. Karena besarnya ukuran data dan banyaknya label dari dataset induk yang mencapai 85 atribut, kami memutuskan untuk memilih 15 atribut yang relevan dalam penelitian kami untuk dimasukkan dianalisis. Berikut ini adalah label-label yang relevan yang akan digunakan dalam penelitian ini.

Tabel 1.1 – Atribut Relevan Dataset

Nama Label	**Deskripsi**	**Tipe**	**Alasan Pemilihan**
Protocol	Jenis protocol jaringan (TCP/UDP/ICMP)	Kategorikal (Nominal)	Indikator jenis lalu lintas tingkat tinggi
Flow Duration	Durasi total koneksi	Numerik (Kontinu)	Ukuran lalu lintas fundamental (kandidat regresi primer)
Total Fwd Packets	Jumlah paket terkirim maju (*forward)*	Numerik (Diskret)	Intensitas aliran pada arah pengiriman
Total Backward Packets	Jumlah paket terkirim mundur (*backward)*	Numerik (Diskret)	Intensitas aliran pada arah penerimaan
Flow Bytes/s	Kecepatan alir (bytes/sec)	Numerik (Kontinu)	Kecepatan volume lalu lintas (dapat melonjak selama serangan)
Flow Packets/s	Kecepatan alir (packets/sec)	Numerik (Kontinu)	Perilaku laju paket (dapat mengidentifikasi DDoS)
Fwd Packet Length Mean	Ukuran rerata (*forward packets*)	Numerik (Kontinu)	Berguna untuk mengkarakterisasi profil lalu lintas
Bwd Packet Length Mean	Ukuran rerata (*backward packets*)	Numerik (Kontinu)	Mirip, tetapi untuk respons
SYN Flag Count	Jumlah SYN (*synchronizes sequence numbers*) *flags*	Numerik (Diskret)	Penting untuk mengidentifikasi serangan banjir SYN
ACK Flag Count	Jumlah ACK (*acknowledgment*) *flags*	Numerik (Diskret)	Indikator pembentukan sesi
Down/Up Ratio	Rasio unduhan terhadap unggahan	Numerik (Kontinu)	Pola perilaku: eksfiltrasi, aliran yang banyak diunduh
Average Packet Size	Ukuran rerata besar paket	Numerik (Kontinu)	Efisiensi ukuran atau petunjuk anomali
Active Mean	Durasi aktif rata-rata selama aliran	Numerik (Kontinu)	Berapa lama koneksi tetap aktif bertukar data
Idle Mean	Durasi idle rata-rata selama aliran	Numerik (Kontinu)	Idle yang lama dapat menunjukkan aktivitas yang mencurigakan/laju rendah
Label	Klasifikasi serangan/jinak	Kategorikal (Nominal)	Variabel target untuk analisis dan visualisasi