ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - TF 181801

Evaluasi Algoritma LSTM dan Algoritma Validasi Sekuensi ID Untuk Mendeteksi Serangan Pada Protokol Komunikasi Modbus TCP/IP Dalam SCADA

I Gede Sanjaya Putra Vhyasa
NRP. 02311940000083

Dosen Pembimbing
Dr. Bambang L. Widjiantoro, S.T., M.T.
NIP.19690507 199512 1 001

Program Studi S1 Teknik Fisika
Departemen Teknik Fisika
Fakultas Teknologi Industri dan Rekayasa Sistem
Institut Teknologi Sepuluh Nopember
Surabaya
2023

Tugas Akhir - TF 181801

Evaluasi Algoritma LSTM dan Validasi Sekuensi ID Untuk Mendeteksi Serangan Pada Protokol Komunikasi Modbus TCP/IP Dalam SCADA

I GEDE SANJAYA PUTRA VHYASA

NRP. 02311940000083

Dosen Pembimbing

Dr. Bambang L. Widjiantoro, S.T., M.T.

NIP 19690507 199512 1 001

Program Studi S1 Teknik Fisika
Departemen Teknik Fisika
Fakultas Teknologi Industri dan Rekayasa Sistem
Institut Teknologi Sepuluh Nopember
Surabaya
2023

Halaman ini sengaja dikosongkan

FINAL PROJECT - TF 181801

Evaluation of LSTM Algorithm and Validation of ID Sequences to Detect Attacks on Modbus TCP/IP Communication Protocol in SCADA

I GEDE SANJAYA PUTRA VHYASA
NRP. 02311940000083

Advisor
Dr. Bambang L. Widjiantoro, S.T., M.T.
NIP. 19690507 199512 1 001

Study Program Bachelor of Engineering Physics
Department of Engineering Physics
Faculty of Industrial Technology and System Engineering
Institut Teknologi Sepuluh Nopember
Surabaya
2023

Halaman ini sengaja dikosongkan

PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini.

Nama Mahasiswa / NRP : I Gede Sanjaya Putra Vhyasa/02311940000083
Departemen / Prodi : Teknik Fisika / S1 Teknik Fisika
Dosen Pembimbing / NIP : Dr. Bambang L. Widjiantoro, S.T., M.T. /19690507 199512 1 001

dengan ini menyatakan bahwa Tugas Akhir dengan judul "*Evaluasi Algoritma LSTM dan Validasi Sekuensi ID Untuk Mendeteksi Serangan Pada Protokol Komunikasi Modbus TCP/IP Dalam SCADA*" adalah hasil karya saya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Institut Teknologi Sepuluh Nopember.

Surabaya, 27 Juli 2023
Mahasiswa,

(I Gede Sanjaya Putra Vhyasa)
NRP. 02311940000083

Halaman ini sengaja dikosongkan

STATEMENT OF ORIGINALITY

The undersigned below:

Name of student / NRP : I Gede Sanjaya Putra Vhyasa/ 02311940000083
Department : Engineering Physics
Advisor / NIP : Dr. Bambang L. Widjiantoro, S.T., M.T. /19690507 199512 1 001

hereby declare that the Final Project with the title of "Evaluation of LSTM Algorithm and Validation of ID Sequences to Detect Attacks on Modbus TCP/IP Communication Protocol in SCADA" is the result of my own work, is original, and is written by following the rules of scientific writing.

If in the future there is a discrepancy with this statement, then I am willing to accept sanctions in accordance with the provisions that apply at Institut Teknologi Sepuluh Nopember.

Surabaya, 27 July 2023
Student

(I Gede Sanjaya Putra Vhyasa)
NRP. 02311940000083

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN
TUGAS AKHIR

Evaluasi Algoritma LSTM dan Validasi Sekuensi ID Untuk Mendeteksi Serangan Pada Protokol Komunikasi Modbus TCP/IP Dalam SCADA

Oleh:

I Gede Sanjaya Putra Vhyasa

NRP. 02311940000083

Surabaya, 27 Juli 2023

Menyetujui,

Pembimbing I

Dr. Bambang L. Widjiantoro, S.T., M.T.

NIP.19690507 199512 1 001

Mengetahui,

Kepala Departemen

Teknik Fisika FT-IRS ITS

Dr. Suyanto, S.T., M.T.

NIP. 19711113 199512 1 002

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

Evaluasi Algoritma LSTM dan Validasi Sekuensi ID Untuk Mendeteksi Serangan Pada Protokol Komunikasi Modbus TCP/IP Dalam SCADA

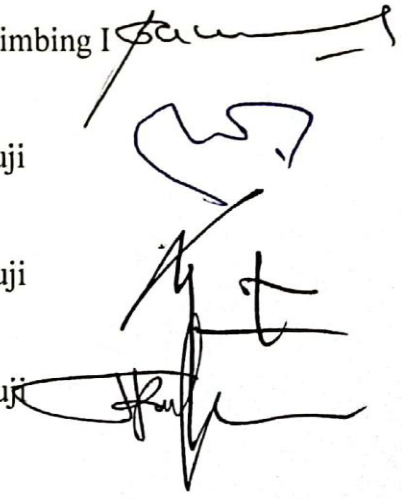
TUGAS AKHIR

Diajukan untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Teknik pada
Program Studi Sarjana Teknik Fisika
Departemen Teknik Fisika
Fakultas Teknologi Industri dan Rekayasa Sistem
Institut Teknologi Sepuluh Nopember

Oleh: **I Gede Sanjaya Putra Vhyasa**
NRP. **0231940000083**

Disetujui oleh Tim Penguji Tugas Akhir:

1. Dr. Bambang L. Widjiantoro, S.T., M.T. — Pembimbing I
2. Andi Rahmadiansah, S.T. M.T. — Penguji
3. Dr. Suyanto, S.T., M.T. — Penguji
4. Iwan Cony Setiadi, S.T., M.T. — Penguji



SURABAYA
Juli, 2023

Halaman ini sengaja dikosongkan

APPROVAL SHEET

Evaluation of LSTM Algorithm and Validation of ID Sequences to Detect Attacks on Modbus TCP/IP Communication Protocol in SCADA

FINAL PROJECT

Submitted to fulfill one of the requirements
for obtaining a degree Bachelor of Engineering at
Undergraduate Study Program of Engineering Physics
Department of Engineering Physics
Faculty of Industrial Technology and Systems Engineering
Institut Teknologi Sepuluh Nopember

by:

I Gede Sanjaya Putra Vhyasa

NRP. 02311940000083

Approved by the Final Assignment Examiner Team:

1. Dr. Bambang L. Widjiantoro, S.T., M.T. — Advisor
2. Andi Rahmadiansah, S.T. M.T. — Examiner
3. Dr. Suyanto, S.T., M.T. — Examiner
4. Iwan Cony Setiadi, S.T., M.T. — Examiner

SURABAYA

July, 2023

Halaman ini sengaja dikosongkan

Evaluasi Algoritma LSTM dan Validasi Sekuensi ID Untuk Mendeteksi Serangan Pada Protokol Komunikasi Modbus TCP/IP Dalam SCADA

Nama Mahasiswa / NRP : I Gede Sanjaya Putra Vhyasa/02311940000083
Departemen : Teknik Fisika FTIRS – ITS
Dosen Pembimbing : Dr. Bambang L. Widjiantoro, S.T., M.T.

Abstrak

Pesatnya perkembangan IoT terutama dengan penerapan teknologi 5G, SCADA menjadi protokol yang semakin banyak diminta yang dulunya hanya dikembangkan di lingkungan yang hampir tidak memerlukan dan menerapkan keamanan, menjadi target utama serangan cyber. Oleh karena itu, implementasi *Intrusion Detection System* (IDS) yang tepat menjadi penting. Diperlukan metode yang dapat mendeteksi penyusup dalam sistem. Metode neural network menjadi metode yang cukup terkenal dalam pendeteksi penyusup dan memiliki hasil yang baik tetapi, sifat neural network yang rumit dan memakan banyak waktu untuk melatih. Metode lainnya adalah metode validasi sekuensi ID yang sebelumnya diusulkan untuk CanBus. Kedua metode ini dievaluasi dalam penelitian ini dan ditemukan bahwa LSTM lebih unggul dengan akurasi 99,7%, presisi 99.74%, recal 99.7%, dan F1-Score 99,69%

Kata Kunci: ICS, SCADA, LSTM, Validasi Sekuensi ID

Halaman ini sengaja dikosongkan

Evaluation of LSTM Algorithm and Validation of ID Sequences to Detect Attacks on Modbus TCP/IP Communication Protocol in SCADA

Student Name /NRP : **I Gede Sanjaya Putra Vhyasa/02311940000083**
Department : **Engineering Physics FTIRS – ITS**
Advisor : **Dr. Bambang L. Widjiantoro, S.T., M.T.**

Abstract

With the rapid development of IoT especially with the adoption of 5G technology, SCADA is becoming an increasingly requested protocol that was previously only developed in environments that required little or no security, becoming a prime target for cyber attacks. Therefore, proper implementation of Intrusion Detection System (IDS) is important. A method is needed that can detect intruders in the system. The neural network method is a well-known method for intruder detection and has good results, however, the nature of the neural network is complicated and takes a lot of time to train. Another method is the previously proposed ID sequence validation method for CanBus. Both of these methods were evaluated in this study and it was found that LSTM was superior with 99.7% accuracy, 99.74% precision, 99.7% recall, and 99.69% F1-Score.

Keywords: ICS, SCADA, LSTM, ID Sequence Validation

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur atas kehadirat Tuhan Yang Maha Esa atas rahmat-Nya sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul: “Evaluasi Algoritma Pendeteksi Penyusup Dengan Metode LSTM dan Validasi Sekuensi ID”. Pada kesempatan kali ini, penulis juga menyampaikan terima kasih kepada :Lorem ipsum dolor sit amet, consetetur sadipscing elitr

1. Bapak Dr. Suyanto, S.T., M.T. selaku Kepala Departemen Teknik Fisika ITS.
2. Bapak Moh. Kamalul Wafi, S.T., MSc.DIC selaku dosen wali yang sudah menemani penulis semasa perkuliahan
3. Bapak Dr. Bambang L. Widjiantoro, S.T., M.T. Sebagai dosen pembimbing
4. Bapak Andi Rahmadiansah, S.T. M.T. Selaku pempmpin penguji dan pengarah kodingan program
5. Keluarga semua yang sangat membantu penulis dalam bentuk fisik dan rohani
6. Teman-teman yang menemani dan membantu penulis dalam pelaksanaan Tugas Akhir

Serta pihak-pihak lain yang tidak dapat disebutkan satu-persatu. Semoga laporan tugas akhir ini dapat dipergunakan dengan sebaik-baiknya.

Surabaya, 27 Juli 2023

I Gede Sanjaya Putra Vhyasa

Halaman ini sengaja dikosongkan

DAFTAR ISI

DAFTAR GAMBAR

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Halaman ini sengaja dikosongkan

BAB I
PENDAHULUAN

1.1 Latar Belakang

Supervisory Control and Data Acquitisition (SCADA) merupakan protocol yang menyambungkan *programable control logic* (PLC), *remote terminal unit* (RTU), *Human-Machine-Interface* (HMI) dan system lainnya. Dengan pesatnya perkembangan IoT terutama dengan penerapan teknologi 5G, SCADA menjadi protokol yang semakin banyak diminta. Dengan tingginya minat SCADA, yang dulunya hanya dikembangkan di lingkungan yang hampir tidak memerlukan dan menerapkan keamanan, menjadi target utama serangan cyber. Oleh karena itu, implementasi *Intrusion Detection System* (IDS) yang tepat menjadi penting.

Sebelum membahas IDS penting diketahui bahwa salah satu protokol SCADA yang paling umum adalah protocol komunikasi Modbus. Awalnya dirancang sebagai komunikasi jalur serial, modbus dibuat dengan asumsi bahwa fungsi SCADA berfungsi secara ideal menyebabkan kebanyakan sistem modbus memiliki sedikit atau tidak ada mekanisme pertahanan terhadap serangan yang disengaja. Protokol tidak memasukkan otentikasi dan otorisasi, dan tidak ada verifikasi integritas data. Selain itu, semua data ditransfer dalam bentuk teks biasa, tanpa enkripsi apa pun [1]

IDS diadopsi secara besar untuk membuat lingkungan cyber aman. IDS juga memainkan peran yang sama dalam SCADA. Terdapat dua jenis IDS *signature-based* dan *anomaly-based* [2]. *Signarure-based* menemukan serangan jaringan melalui ekstraksi sidik jari dari paket data dan mengidentifikasinya. Contohnya pada [3], V. Jan & H. Martin, mengevaluasi aturan Snort dan Quickdraw berdasarkan tanda tangan untuk menentukan hubungannya dengan keamanan cyber SCADA. ”Aturan” dalam snort adalah komponen fundamental yang digunakan untuk deteksi dan pencegahan intrusi jaringan. Aturan Snort adalah ekspresi berbasis teks yang menentukan kondisi khusus untuk mengidentifikasi pola lalu lintas jaringan yang terkait dengan potensi ancaman keamanan, sedangkan quickdraw, aturan tidak ditentukan secara eksplisit. Sebaliknya, gim ini mengandalkan model jaringan saraf convolutional (CNN) terlatih untuk mengenali dan mengklasifikasikan sketsa yang digambar tangan.

Anomaly-based detection mengamati serangan itu pola dari data melalui algoritma pembelajaran mesin. Pada [4], A. Simon Duque et al. Mengevaluasi machine learning untuk mendeteksi deteksi anomaly pada dataset protocol modbus TCP/IP dan menemukan bahwa

machine learning dapat mendeteksi anomaly secara baik. Mengetahui ini penggunaan Deep Learning, machine learning yang lebih rumit, mulai digunakan dalam pendeteksi anomaly pada data modbus. Pada penelitian ini, Deep Learning LSTM *neural network* akan digunakan untuk mendeteksi adanya anomaly pada dataset Electra Modbus. Metode ini akan dibandingkan dengan metode yang diusulkan oleh Marchetti et al [5] dimana mereka menggunakan transisi matriks berisikan data CanBus sah dan membandingkannya dengan data yang memiliki anomali. Penelitian ini juga akan melihat tipe serangan apa saja yang mudah untuk dideteksi, dan fitur pada dataset apa saja yang mempengaruhi performa kedua metode tersebut.

1.2 Rumusan Masalah

Permasalahan umum yang ingin dikaji berdasarkan latar belakang di atas adalah:

a) Bagaimana hasil evaluasi metode LSTM dan metode validasi ID?
b) Tipe serangan apa saja yang mudah dideteksi oleh kedua metode?
c) Bisakah hasil akurasi, presisi, recal, dan skor F1 metode validasi ID bersaing dengan metode LSTM?

1.3 Tujuan

Tujuan dari penelitian ini adalah:

a) Mengetahui hasil akurasi, presisi, recal, dan skor F1 metode LSTM dan metode Validasi Sekuensi ID
b) Mendeteksi serangan apa saja yang dapat menembus deteksi kedua metode
c) Membandingkan performa kedua metode berdasarkan hasil akurasi, presisi, recal, dan skor F1

1.4 Batasan Masalah

Penelitian menekankan beberapa batasan masalah, yaitu:

a) Protokol komunikasi yang digunakan adalah Modbus TCP/IP
b) Data yang digunakan merupakan data sekunder komunikasi Modbus
c) Metode deteksi masalah menggunakan machine learning dan transisi matriks
d) Data memiliki ID unik yang berulang

1.5 Sistematika Laporan

Laporan ini akan membuka dengan menyebutkan betapa pentingnya SCADA dalam ICS dan protokol komunikasi apa yang umum berada pada sistem SCADA. Pendahuluan juga akan membahas mengenai kerentanan pada sistem SCADA terutrama di protokol komunikasinya dan menyebutkan cara mengatasi ini. Selain itu disebutkan juga rumusan dan tujuan dari penelitian ini. Pada bab dua akan dijelaskan lebih mendalam mengenai SCADA, protokol ModBus, dataset yang digunakan, kedua metode yang akan digunakan dan serangan yang dapat memmpengaruhi SCADA serta protokol komunikasi Modbus. Laporan akan menjelaskan alur penelitian pada bab tiga dan alur algoritma pada kedua metode. Hasil dari penelitian ini akan diperlihatkan di bab empat dan akan disimpulkan di bab lima.

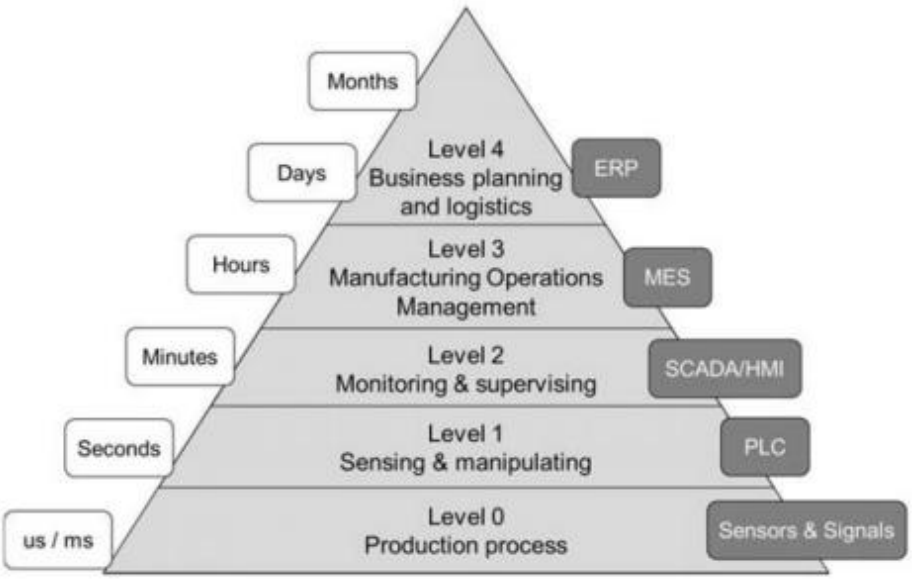
Halaman ini sengaja dikosongkan

BAB II
TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 SCADA Dalam Automasi Industri

Dalam definisinya, National Institute of Standard and Technology (NIST), mendeskripsikan bahwa ICS adalah istilah umum yang mencakup system kontrol yang berbeda seperti sistem Supervisory Control and Data (SCADA), Programing Logic Control (PLC) dan konfigurasi system lainnya [6].

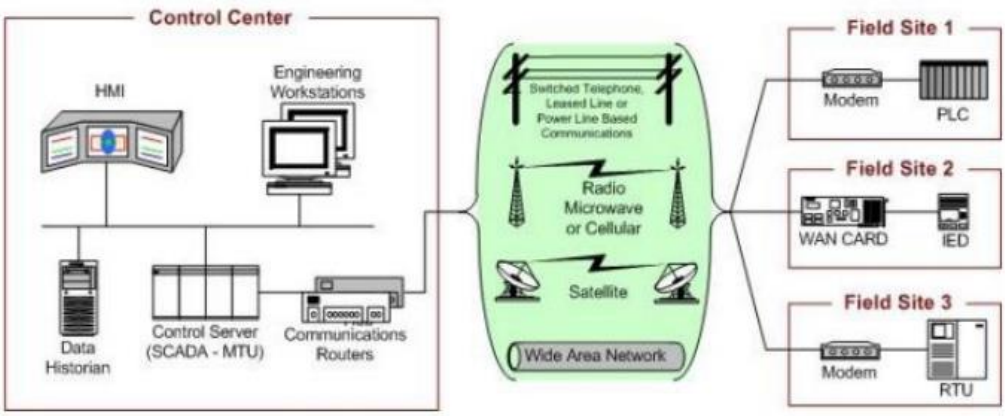
ICS sendiri memiliki lapisan automasi yang dapat dijelaskan dalam piramid automasi. Piramida otomasi mengklasifikasikan berbagai lapisan teknologi informasi dari pabrik produksi. Contoh piramida automasi dapat dilihat pada gambar 2.1



Gambar 2.1 Piramida Otomasi. Sumber: [6]

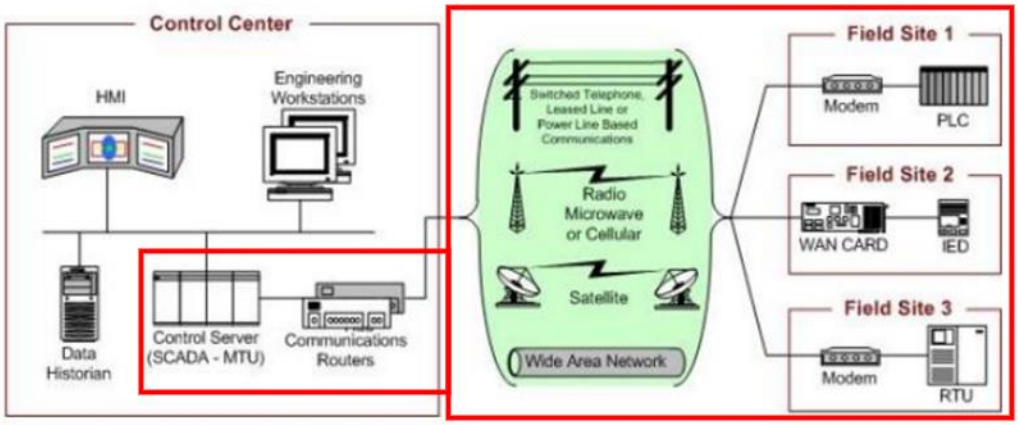
Sistem SCADA, biasanya berada pada bagian kedua. SCADA memungkinkan komponen otomasi elektronik untuk berkomunikasi meskipun tersebar secara geografis. SCADA secara dasar terdiri dari pusat kendali dan beberapa lokasi lapangan yang terhubung melalui Wide Area Network (WAN). Dalam dunia otomasi industri, sistem SCADA digunakan untuk memantau dan mengontrol distribusi air, minyak dan gas alam, termasuk jaringan pipa, kapal, truk, dan sistem kereta api, serta sistem pengumpulan air limbah.

Komunikasi antar alat biasanya terjadi melalui poling dan protokol komunikasi modbus, server mencatat data dari lokasi lapangan dan mengirimkan inforamasi ke Human-Machine-Interface (HMI) seperti yang terlihat di gambar 2.2.



Gambar 2.2 Arsitektur Umum SCADA. Sumber: [7]

Ancaman keamanan sistem SCADA mendapatkan perhatian yang tinggi dari peneliti dan pengembang karena sistem SCADA semakin rentan seiring banyaknya konektivitas dan kompleksitas [8] kerentanan ini dikarenakan, awalnya, sistem SCADA dirancang untuk digunakan dalam jaringan terisolasi dan menghubungkannya ke jaringan lain, seperti Internet, membuatnya jauh lebih rentan [9]. Mengetahui ini, dapat dilihat gambaran kerentanan SISTEM SCADA berada pada koektivitas SCADA itu sendiri, jika gambar 2.2 merupakan aristektur umum SCADA, maka gambar 2.3 adalah bagian yang rentan terhadap serangan.



Gambar 2.3 Kerentanan Pada Sistem SCADA

Pada gambar 2.3 merupakan bagian yang rentan terhadap serangan. Serangan dapat mempengaruhi server yang akan mempengaruhi peralatan lapangan jika dibiarkan. Beberapa serangan yang umum terjadi adalah Denial of Service, Bombs (Logic or Time), Masquerade

Serangan Replay, Intercept/Alter, Service Spoofing, Substitution, Scavenging, Tunneling, Unauthorized Access, Trap Door/ Back Door. Beberapa serangan ini mengandalkan dan/atau meneksploitasi protokol komunikasi yang sering digunakan dalam SCADA, Modbus.

2.2 Pengenalan Modbus

Komunikasi umum yang digunakan dalam sistem SCADA adalah ModBus. Sebelumnya diketahui sebagai modicon, Modbus dibuat pada tahun 1979 oleh Schneider-Electric. Modbus awalnya merupakan protocol komunikasi serial antara Remote Terminal Unit (RTU) dan PLCs [10]. Sejak saat itu, protokol komunikasi modbus sering digunakan oleh perusahaan indsutri sebagai standar protokol komunikasi antar alat [11]. Dengan keuntungan seperti bebas royalti dan relatif mudah untuk diemplementasikan tidak aneh bahwa protokol komunikasi ini menjadi sangat populer dalam industri otomasi.

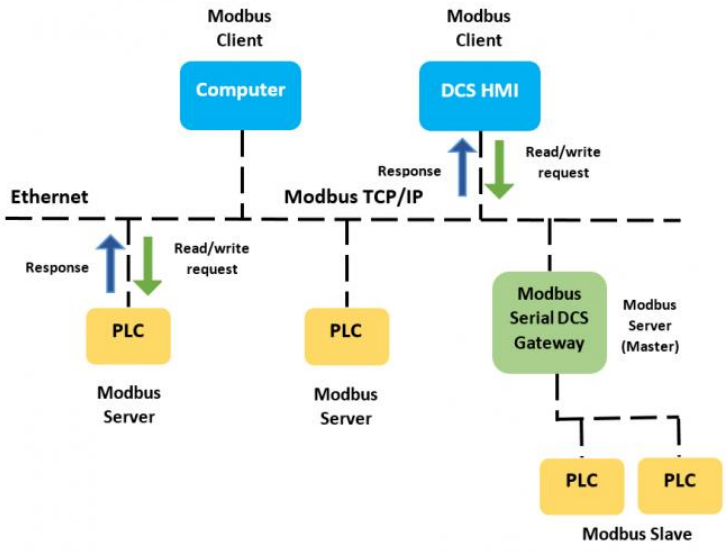
Sekarang modbus memiliki banyak versi. Versi yang banyak digunakan tercantum pada tabel 2.1

Tabel 2.1
Tipe-Tipe Modbus dan Deskripsi Singkatnya

Versi	Deskripsi
Modbus RTU	Digunakan dalam komunikasi serial & menggunakan representasi nilai data biner yang dipadatkan
Modbus ASCII	Protokol komunikasi serial yan memanfaatkan karakter ASCII.
Modbus TCP/IP	Berkomunikasi dengan protokol stack TCP/IP

Selama bertahun-tahun, tren ekonomi mendorong teknologi modbusdari sistem komunikasi serial kecil ke jaringan berskala besar berdasarkan TCP/IP. Hal ini, membuat modbus TCP/IP menjadi pilihan yang menjadi populer dalam otomasi industri. Dalam modbus TCP/IP, data dienkapsulasi dalam paket TCP/IP dan ditransmisikan pada frame ethernet [12]. Perbedaan signifikan antara komunikasi Modbus default ke Modbus/TCP adalah bahwa Modbus/TCP menggunakan arsitektur klien/server melalui Ethernet. Modbus yang khas

Arsitektur TCP yang terdiri dari komunikasi Modbus serial dan komunikasi Modbus/TCP secara umum dapat dilihat pada gambar 2.4



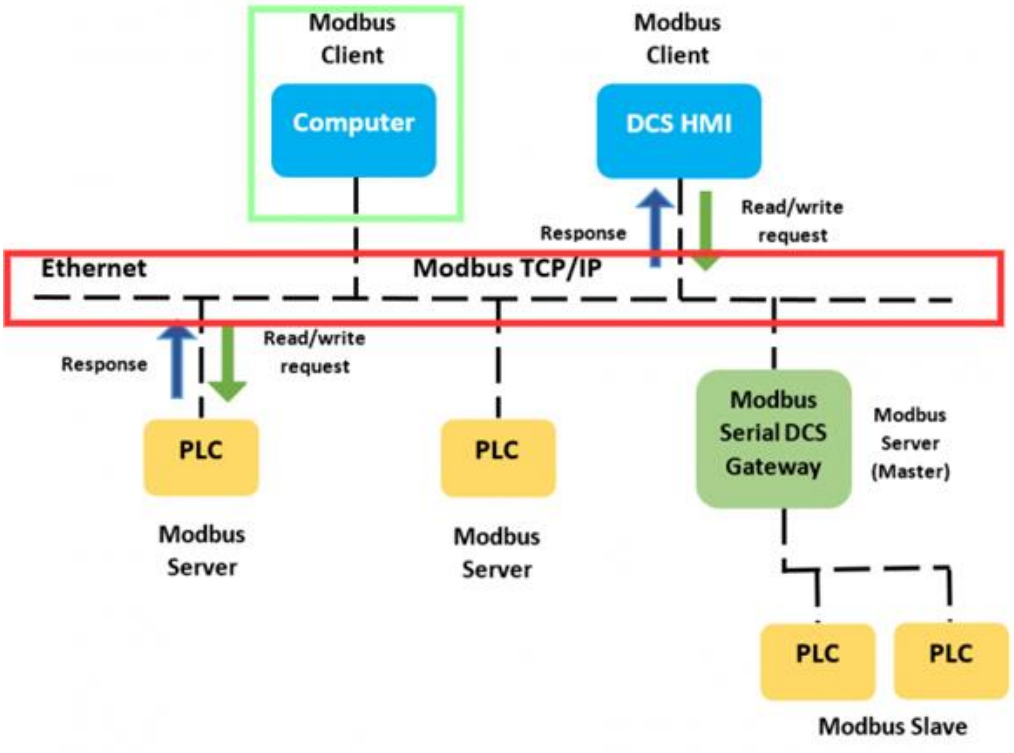
Gambar 2.4 Arsitektur Umum Modbus TCP/IP. Sumber: [8]

Pada gambar 2.4 dapat diambil bagaimana protokol komunikasi modbus tcp/ip bekerja. Data dari peralatan yang berada di lapangan dibungkus dalam header dan dikirim ke HMI yang berperan sebagai client dari modbus tcp/ip. Client dalam modbus tcp/ip dapat mengirim kueri kepada server, dan server akan mengirimkan apa yang diminta oleh client.

Penggunaan protokol ethernet atau IP pada Modbus TCP/IP membuat Modbus menjadi rentan terhadap serangan. pada jurnalnya [14] Morris et al, mengklasifikasikan serangan terhadap industri modbus kedalam tujih kategori. Serangan pengintaian mengumpulkan informasi jaringan sistem kontrol, *Naive Malicious Response Injection(NMRI)* serangan injeksi paket dengan kurangnya kontrol dan monitor, *Complex Malicious Response Injection* mirip dengan NMRI tetapi injeksi ini berupaya untuk menyamarkan keadaan sebenarnya untuk memengaruhi, secara negatif, loop kontrol umpan balik yang mengelola sistem fisik siber. *Malicious State Command Injection* (MSCI) mengubah keadaan sistem kontrol dari keadaan aman menjadi keadaan kritis dengan menyuntikkan perintah jahat ke perangkat lapangan, *Malicious Parameter Command Injection*(MPCI) menginjeksikan perintah yang dapat mengganti setpoint pada kontrol, *Malicious Function Code Injection* (MFCI) menginjeksikan kode fungsi untuk mengubah perintah modbus contoh, jika modbus melakukan write/read, serangan akan mengganti menjadi read saja. Terakhir adalah *Denial of Service* (DOS) yaitu serangan dengan tujuan menghentikan berfungsinya beberapa bagian sistem fisik dunia maya untuk menonaktifkan seluruh sistem secara efektif.

2.3 Penerapan IDS pada Modbus

Kerentanan pada sistem SCADA dan protokol komunikasi Modbus menyebabkan perhatian yang meningkat ke sistem keamanan pada SCADA dan Modbus. Deteksi serangan dapat dikategorikan menjadi dua, *signature-based* dan *anomaly-based*. Pengimplementasian IDS bervariasi berdasarkan sumber metrik yang dikumpulkan dari sistem [8]. Hal yang dapat dipantau adalah jaringan, host, aplikasi, atau metrik kritis. Barbosa, et al [11] menggunakan model flow dalam *anomaly*-based IDS untuk menggambarkan lalu lintas jaringan dengan harapan bahwa IDS dapat mendeteksi adanya kejanggalan. Usul dari Barbosa, et al dapat dijadikan dasar deteksi serangan dengan memonitor lalu lintas data dan menggunakan kedua metode yang ingin dievaluasi dalam penelitian ini. Data lalu lintas jaringan dapat dimonitor dalam HMI dengan saplikasi yang tersedia, seperti Wireshark. Data ini kemudian dimasukkan ke dalam algoritma dan ditentukan apakah lalu lintas aman atau memiliki serangan. Jika gambar 2.4 adalah arsitektur modbus yang digunakan, gambar 2.5 adalah gambaran dimana IDS dapat diimplementasikan.



Gambar 2.5 Pengimplementasian deteksi serangan dengan metode yang mengandalkan lalu lintas data Modbus

Garis merah merupakan bagian modbus yang rentan terhadap serangan dan yang dimonitor dan garis hijau adalah tempat memonitor lalu lintas jaringan. Algoritma pendeteksi serangan dapat diimplementasikan pada HMI yang memonitor jaringan komunikasi data karena

dalam deteksi berbasis anomali mendeteksi sifat abnormal dari ICS dan/atau lalu lintas jaringan [8]. Log dari lalu lintas dapat digunakan sebagai input algoritma pendeteksi, dan algoritma pendeteksi dapat menentukan apakah lalu lintas aman, atau tidak. Jika algoritma menaikkan *flag* maka akan ada pemberitahuan dan/atau alarm yang menunjukan adanya serangan pada sistem

2.4 Dataset

Seperti yang telah disebutkan sebelumnya, peralihan modbus dari komunikasi serial kecil ke jaringan skala besar dengan protokol TCP/IP, konektivitas TCP/IP memaparkan sistem modbus yang sebelumnya tertutup terhadap serangan jaringan jarak jauh, bahkan melalui internet. Peningkatan serangan pada ICS mempengaruhi penggunaan *Intrusion Detection System* (IDS) di bidang ICS. Sejumlah besar teknik deteksi intrusi telah diusulkan dalam literatur untuk mengatasi ancaman ini. Saat ini, teknik yang memiliki kinerja yang baik adalah Machine Learning dan Deep Learning [15], [16]. Deteksi penyusup juga dapat dilakukan dengan melihat ID yang berulang agar bisa melihat kejanggalan dalam sistem [5].

Metode-metode ini diukur performanya menggunakan dataset yang berisi data relevan (lalu lintas jaringan, log sensor dan aktuator, atau fitur dari sumber sebelumnya) dari skenario ICS di mana beberapa serangan sedang berjalan. Banyak dataset yang telah dibuat sontohnya adalah dataset yang dibuat oleh Tiexeira et al. Mereka mengembangkan sistem SCADA testbed yang terdiri dari sistem kontrol tangki penyimpanan air untuk pengolahan dan distribusi air. Data dikumpulkan dengan perangkat lunak WireShark untuk mendapatkan file PCAP. Penulis melakukan lima jenis serangan pengintaian yang berbeda dalam testbed mereka. Namun, kumpulan data masih kurang karena tipe serangan hanya terdiri dari serangan pengintaian [17]. Rosa, et al. Mengembangkan testbed sistem SCADA dalam lingkungan *hybrid* yang terdiri dari komunikasi jaringan asli dan aset SCADA yang meniru jaringan listrik [18]. Dataset ini memiliki tipe serangan yang komprehensif. Selain itu testbed meniru jaringan listrik, hal ini merupakan hal yang penting dalam kehidupan sehari-hari dan merupakan industri yang populer. Sayangnya kumpulan data tidak dapat dibagikan di internet ataupun sumber terbuka lainnya untuk dianalisis.

Dataset yang diambil pada penelitian ini adalah electra modbus dataset. Electra dataset yang menirukan gardu listrik kereta api berkecapatan tinggi. Tujuan utama dari testbed ini adalah untuk memungkinkan konversi daya listrik jaringan umum menjadi kondisi tegangan,

arus, dan frekuensi untuk memasok kereta api atau trem. Pada dataset ini terdapat tiga kategori serangan yaitu serangan pengintaian, serangan injeksi data salah, dan serangan *replay* yang membuat dataset ini baik untuk dijadikan untuk menguji performa algoritma pendeteksi penyusup yang diusulkan [19].

2.5 Long-Short Term Memory

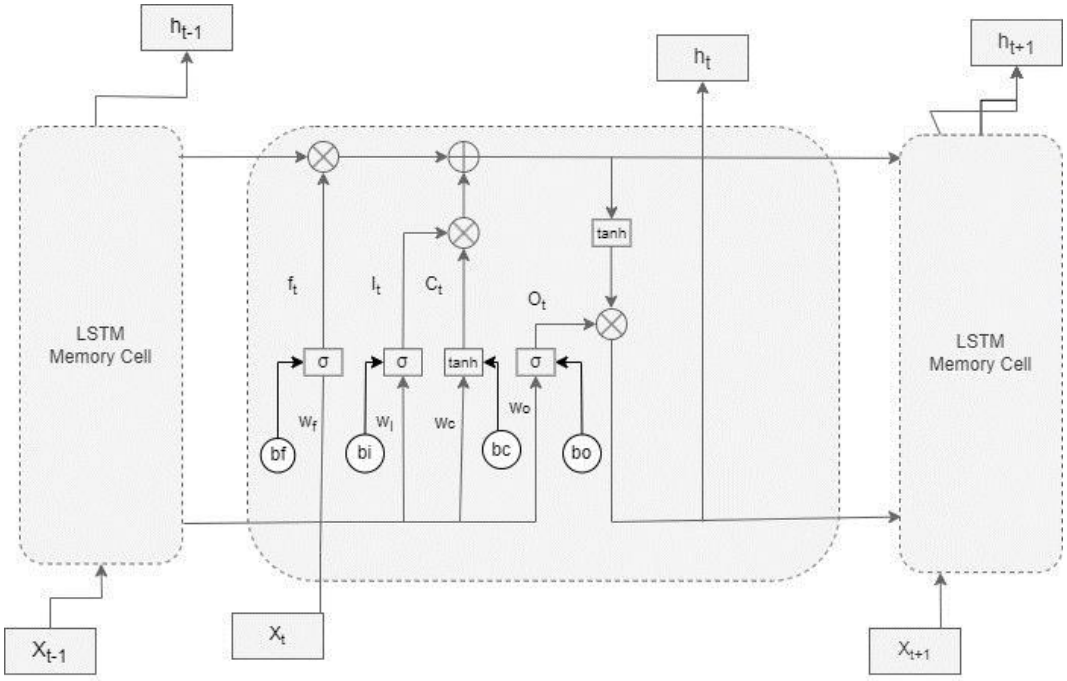
Sebelumnya telah diungkit bahwa Deep Learning megnhasilkan deteksi penyusup yang baik. Salah satu Deep Learning yang akan digunakan pada penelitian ini adalah *Long-Short Term Memory*. *Long-Short Term Memory* merupakan merupakan *Recurrent Neural Network* bertujuan untuk menangani masalah gradien menghilang yang ada di RNN tradisional. Seperti RNN, LSTM juga merupakan algoritma yang mengandalkan sequensi dalam data yang dimasukkan ke algoritma untuk memprediksi data selanjutnya. Beberapa keuntungan dari LSTM adalah:

1. Dapat menangani variable urutan panjang. Hal ini dikarenakan LSTM tidak menganggap ukuran input tetap.
2. Sel memori dalam LSTM memungkinkan informasi bertahan dalam waktu yang lama.
3. LSTM dapat memproses beberapa deret waktu paralel secara bersamaan, memungkinkannya memodelkan ketergantungan dan interaksi antara urutan data yang berbeda.
4. LSTM efektif dalam mempelajari fitur yang berguna dari data berurutan melalui kemampuan pembelajaran representasi hierarkisnya [20].

Pada penelitiannya, F. Rui et al. menggunakan LSTM neural network dan membandingkannya dengan ARIMA dan menyimpulkan bahwa LSTM lebih unggul daripada ARIMA. *Mean Squared Error* ARIMA dan LSTM adalah 841.0065 dan 710.0502 [21].

H. Md Delwar et al. menggunakan LSTM untuk mendeteksi intrusi dalam dataset NAIST CAN attack mendapatkan hasil bahwa LSTM dapat mendeteksi DoS dan serangan Spoofing dengan tingkat deteksi 1,00, dan tingkat deteksi Fuzzing 0,9994 [22].

Arsitektur LSTM yang digunakan pada penelitian ini adalah LSTM vanilla dengan forget gate yang dapat memutuskan apakah sebuah sekuensi diingat atau dilupakan. Berikut adalah gambar dari arsitektur umum LSTM



Gambar 2.6 Diagram Arsitektur LSTM Vanilla. Sumber: [9]

Pada diagram arsitektur diatas, didapatkan bahwa terdapat dua aktivasi sigmoid ($\sigma(x) = \frac{1}{1+e^x}$) dan hyperbolic tangen ($\tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$). Fungsi aktivasi akan mempengaruhi *hidden state* dalam algoritma yang dikomputasi dengan rumus berikut:

$$i_t = \sigma(x_t U^i + h_{t-1} W^i + bi) \quad (1)$$

$$f_t = \sigma\left(x_t U^i + h_{t-1} W^f + \text{bf}\right) \quad (2)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o + \text{bo}) \quad (3)$$

$$C'_t = \tanh(x_t U^g + h_{t-1} W^g + \text{bc}) \quad (4)$$

$$C_t = \sigma(f_t * C_{t-1} + i_t * C'_t) \quad (5)$$

$$h_t = \tanh(C_t) * o_t \quad (6)$$

Pada algoritma LSTM, terdapat empat gate:

1. Cell State (Ct): Status sel mewakili memori LSTM. Ini berjalan secara horizontal di seluruh jaringan, memungkinkan informasi mengalir dari satu langkah waktu ke langkah lainnya. Itu bisa dilihat sebagai ban berjalan yang membawa informasi melalui LSTM. Keadaan sel dapat dimodifikasi melalui penggunaan gerbang.
2. Input Gate (i_t): Gerbang input mengontrol aliran informasi ke dalam keadaan sel. Dibutuhkan input saat ini (x_t) dan status tersembunyi sebelumnya (h_{t-1}), dan menerapkan fungsi aktivasi sigmoid untuk menentukan bagian mana dari input yang harus

disimpan dalam status sel. Keluaran dari gerbang input (i_t) berkisar antara 0 dan 1, mewakili jumlah informasi yang akan dilewatkan.

3. Forgot Gerbang (f_t): Lupakan gerbang menentukan bagian mana dari keadaan sel yang harus dilupakan atau dibuang. Dibutuhkan, sebagai input, input saat ini (x_t) dan status tersembunyi sebelumnya (h_{t-1}), dan menerapkan fungsi aktivasi sigmoid untuk menghasilkan output forgot gate (f_t) antara 0 dan 1. Nilai yang mendekati 1 berarti informasi yang sesuai dalam keadaan sel harus dipertahankan, sedangkan nilai yang mendekati 0 berarti harus dilupakan.
4. Gerbang Keluaran (o_t): Gerbang keluaran mengontrol aliran informasi dari keadaan sel ke keluaran dan keadaan tersembunyi berikutnya. Dibutuhkan input saat ini (x_t) dan status tersembunyi sebelumnya (h_{t-1}), bersama dengan status sel yang diperbarui (C_t), dan menerapkan fungsi aktivasi sigmoid untuk menentukan bagian mana dari status sel yang harus dikeluarkan. Itu juga menerapkan fungsi aktivasi tanh ke keadaan sel saat ini, menghasilkan keadaan tersembunyi baru (h_t) yang membawa informasi ke langkah waktu berikutnya.

Komponen-komponen ini bekerja bersama secara berurutan untuk memproses data input dari waktu ke waktu, memungkinkan LSTM untuk menangkap ketergantungan jangka panjang dan menyimpan informasi untuk waktu yang lebih lama.

2.6 Metode Validasi Sekuensi ID

Terlepas dari kelebihannya, *Neural Network* LSTM masih termasuk lambat karena paralisme yang besar dan sifatnya yang berurutan [10]. M, Marchetti dan S, Dario mengusulkan algoritma yang mengidentifikasikan mengidentifikasi anomali dalam urutan pesan yang mengalir di bus CAN dan ditandai dengan memori kecil dan jejak komputasi, yang membuatnya berlaku untuk ECU saat ini [5].

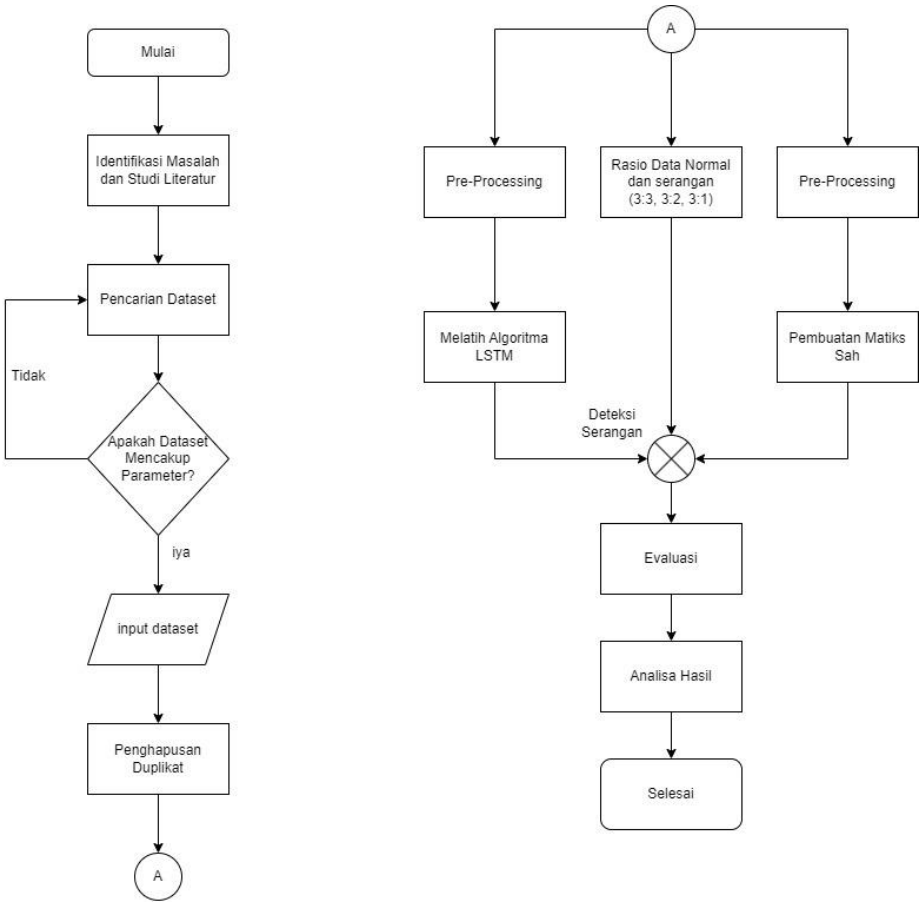
Algoritmanya terbagi menjadi dua fase. Fase pertama adalah fase latihan dimana message CANBus yang normal dijadikan sebagai matriks transisi yang sah dan akan dijadikan sebagai referensi untuk mendeteksi anomali pada fase deteksi. Hasil penelitian menunjukan bahwa algoritma memiliki performa yang baik untuk serangan yang lebih lama dan terdiri dari lebih dari satu pesan. Secara khusus, penelitian menyoroti bahwa urutan yang disusun oleh setidaknya dua pesan selalu terdeteksi (probabilitas deteksi sama dengan 100%), secara independen oleh distribusi probabilitas dari pesan yang disuntikkan..

Halaman ini sengaja dikosongkan

BAB III
METODOLOGI PENELITIAN

3.1 Diagram Alir Penelitian

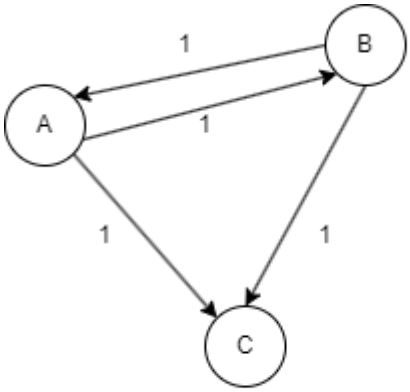
Peneletian ini mengandalakan data sekunder untuk melakukan evaluasi pada dua metode yang telah disebutkan pada bab dua. Gambar 3.1 menggambarkan alur dari awal sampai akhir penelitian



Gambar 3.1 Alur Penelitian

3.2 Metode Validasi ID pada Dataset

Metode yang diusulkan menggunakan matriks transisi untuk memeriksa apakah ada ID yang janggal dalam lalu lintas data. Ada dua fase dalam metode ini fase pembuatan matriks transisi sah dan fase deteksi. Karena metode ini metode ini memanfaatkan matriks transisi untuk membuat matriks sah dan proses validasi, maka gambar 3.2 dapat dijadikan gambaran bagaimana matriks transisi dibuat.



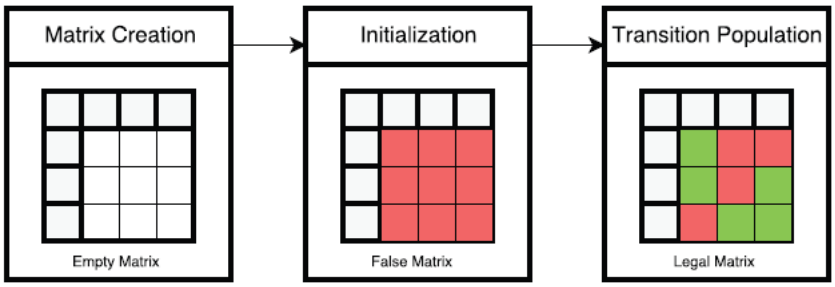
	A	B	C
A	0	1	1
B	1	0	1
C	0	0	

Gambar 3.2 Metode pembuatan matriks transisi

Misalkan ketiga lingkaran adalah baris dari data input yang digunakan. Huruf didalam lingkaran merupakan ID dari baris input. Jika baris i dengan ID sama dengan A, diikuti oleh baris j dengan ID sama dengan B maka value dari baris A dan kolom B akan menjadi 1 atau True. Metode ini akan menghasilkan matriks transisi sah yang menjadi referensi dari data lalu lintas, yang sekuensinya dibandingkan dengan matriks sah

3.2.1 Fase Pembuatan Matriks Sah

Matriks transisi adalah matriks bujur sangkar dengan urutan n, di mana n adalah jumlah ID unik yang tersedia di jejak. Pada awal fase pelatihan, semua nilai matriks transisi diinisialisasi menjadi false. Langkah pertama untuk pembuatan matriks transisi adalah memeriksa urutan ID pada jejak yang sah dan menandai sebagai benar semua transisi yang diamati antara ID. Algoritma untuk populasi matriks transisi direpresentasikan dalam Gambar 3.3

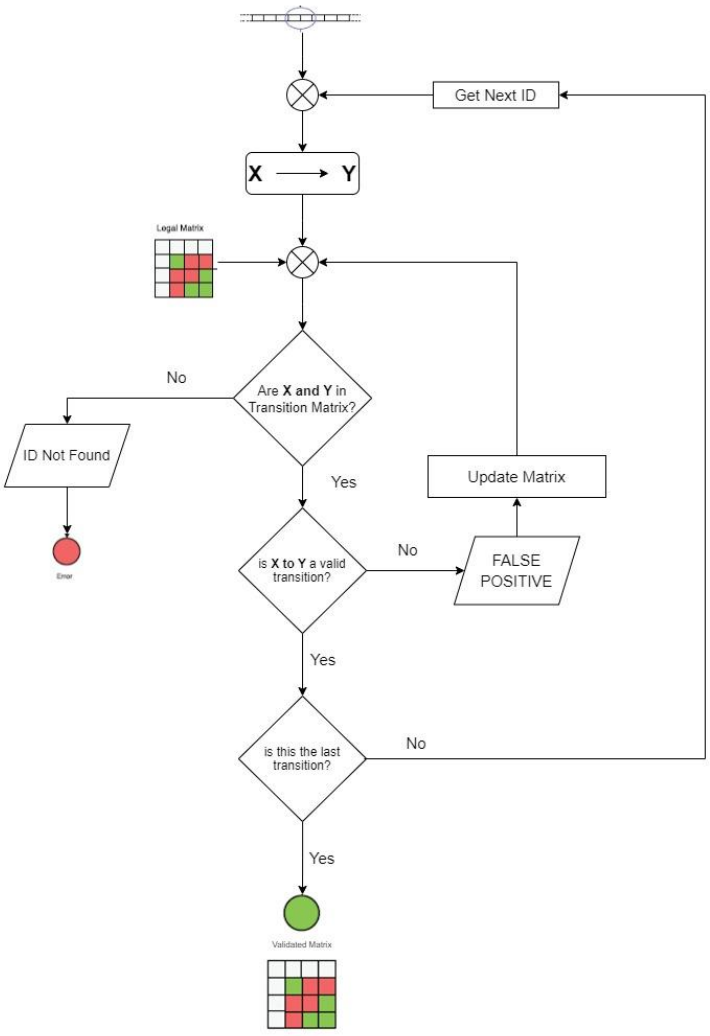


Gambar 3.3 Populasi Matriks Transisi Sah

Struktur data yang dihasilkan berupa matriks yang disusun oleh baik nilai benar atau salah. Struktur data tersebut memungkinkan algoritma final untuk memiliki latensi yang sangat rendah dalam mengakses nilai transisi.

3.2.2 Fase Deteksi

Setelah pembuatan matriks transisi, digunakan jejak lalu lintas resmi lainnya untuk memvalidasi model. Proses validasi mengambil input Matriks Transisi yang dibuat pada langkah sebelumnya, dan urutan dari ID yang diekstraksi dari kumpulan pesan digunakan untuk tujuan validasi. Alur dari deteksi serangan metode validasi sekuensi ID dapat dilihat pada gambar 3.4



Gambar 3.4 Flowchart algoritma validasi sekuensi ID

\Matriks transisi digunakan untuk menguji kemungkinan transisi antara dua ID yang berbeda: jika transisi ditandai sebagai salah dalam matriks, transisi tersebut dianggap sebagai anomali, jika tidak, transisi tersebut sah

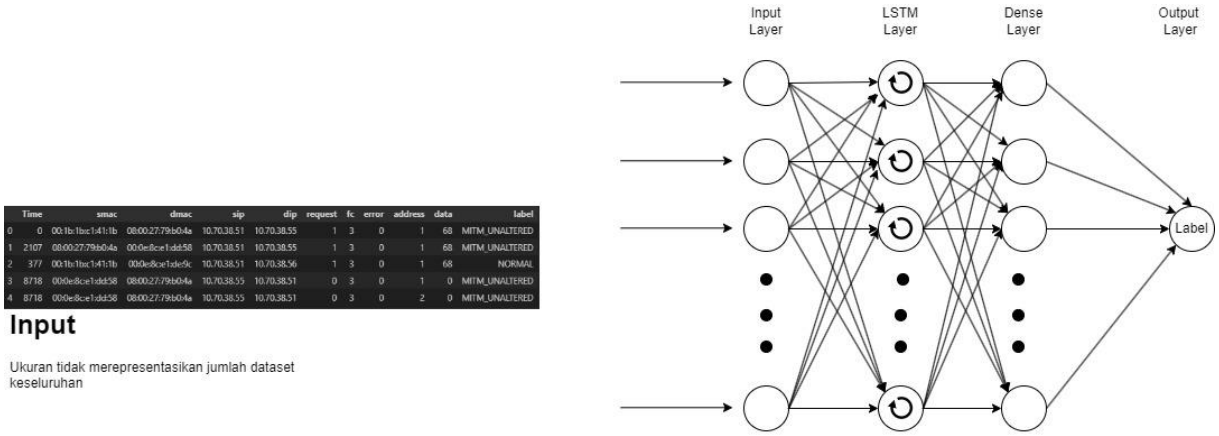
Setelah mengetahui bagaimana algoritma metode ini bekerja maka akan ditentukan fitur apa yang akan dijadikan sebagai ID dalam proses validasi. Empat fitur dipilih dan dibandingkan, evaluasi yang terbaik akan dijadikan sebagai ID validasi. Kemudian tiap serangan dengan jumlah yang berbeda akan digunakan untuk menguji metode ini untuk mengetahui serangan apa yang mudah dan sulit untuk dideteksi.

3.3 Deteksi Dengan Algoritma LSTM

Metode kedua adalah metode pendeteksi penyusup melalui kejanggalan dalam transmisi data dengan menggunakan Long-Short Term Memory. Metode ini dipiih karena LSTM merupakan jenis jaringan saraf berulang (RNN) yang dirancang khusus untuk memproses data berurutan atau deret waktu. Algoritma LSTM juga memiliki dua fase, fase *training* dimana *weight* dan bias akan dioptimalkan dan fase deteksi .Algoritma LSTM disediakan oleh library Keras dalam python.

3.3.1 Fase Training

LSTM akan dilatih dengan variasi hidden layer sebanyak 128, 64, dan 32 dengan tujuan melakukan ablasi pada hidden layer LSTM untuk mengevaluasi variasi mana yang memberikan hasil maksimal. Sebelum ablasi dilakukan, model pertama-tama harus dibuat dan dilatih terlebih dahulu agar mencapai deteksi serangan yang optimal. Gambar 3.5 adalah model neural node dari pembuatan dan pelatihan algoritma LSTM



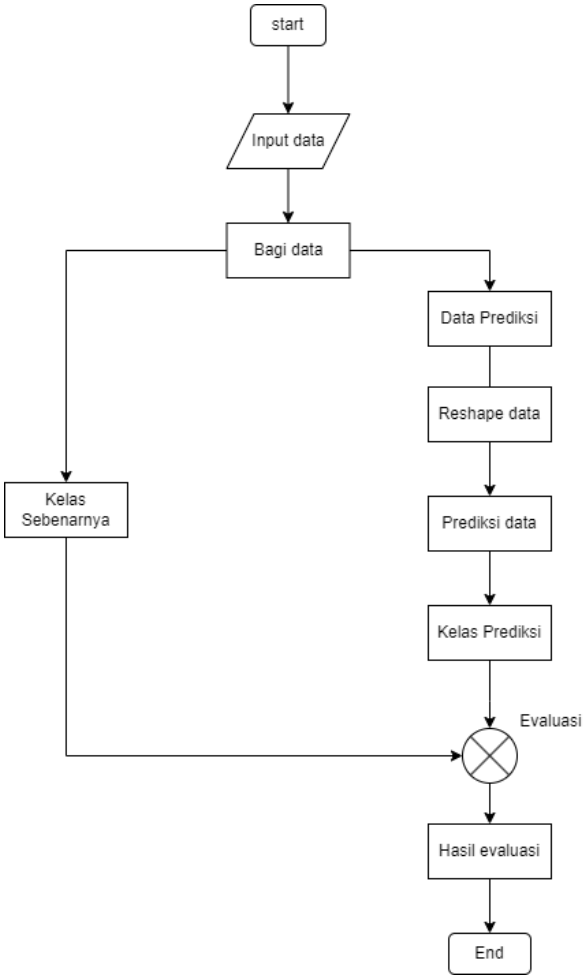
Gambar 3.5 Node LSTM pada Penelitian. Sumber: [14]

Nilai weight dan bias dioptimisasi dengan algoritma pembelajaran Adaptive Moment Estimation (ADAM). Optimisasi Adam, disediakan oleh library keras dalam python,

merupakan optimisasi gabungan dari RMSprop dan Stochastic Gradiendt Descent dengan momentum [14]. Hasil dari optimisasi weight dan bias akan dimasukkan ke dalam lapisan Dense dengan aktivasi *softmax* agar dapat menghasilkan probabilitas tiap kelas dalam output. Model dilatih dengan pengulangan 10 kali

3.3.2 Fase Deteksi

Pada fase deteksi terlebih dahulu data input akan dibagi kedalam data test dan data kelas sebenarnya. Data juga akan direshape sesuai dengan dimensi yang diperlukan oleh model LSTM (jumlah sampel, panjang sequence, jumlah fitur). Selanjutnya, model yang telah dilatih dievaluasi menggunakan data testing. Setelah mendapatkan kelas prediksi dan kelas sebenarnya, metrik evaluasi seperti akurasi), presisi, recall, dan F1-score dihitung. Metrik evaluasi ini memberikan gambaran tentang kinerja model dalam memprediksi data testing.



Gambar 3.6 Flowchart deteksi serangan dengan LSTM

Gambar 3.6 menunjukan proses deteksi serangan pada algoritma LSTM. Langkah ini juga dilakukan dalam ablasi LSTM dengan lapisan 128, 64, dan 32. Tiap hasil evaluasi akan dibandingkan dan dilihat hasil masing-masing. Hasil yang terbaik akan dipilih sebagai model utama untuk melakukan deteksi serangan dengan rasio yang berbeda.

3.4 Pengujian Performa Kedua Algoritma

Setelah melaukan semua langkah di atas, perlu dilakukan pemeriksaan perform algoritma. Kedua algoritma dapat dibandingkan dengan memeriksa kelas hasil prediksi dan kelas yang nyata. Dengan membandinginkan kedua kelas ini didapatkan:

- Hitung True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN):
- TP: Jumlah anomali yang diprediksi dengan benar (diprediksi sebagai 1, label sebenarnya adalah 1).
- TN: Jumlah instance normal yang diprediksi dengan benar (diprediksi sebagai 1, label sebenarnya adalah 1).
- FP: Jumlah anomali yang diprediksi salah (diprediksi sebagai 1, label sebenarnya adalah 2).
- FN: Jumlah instance normal yang diprediksi salah (diprediksi sebagai 2, label sebenarnya adalah 1).

Hitung metrik evaluasi:

- Akurasi: (TP + TN) / (TP + TN + FP + FN)
- Presisi: TP / (TP + FP)
- Recall (Sensitivitas atau Tingkat Positif Sejati): TP / (TP + FN)
- Spesifisitas (Tingkat Negatif Sejati): TN / (TN + FP)
- F1-score: 2 * (Precision * Recall) / (Precision + Recall)

BAB IV
HASIL DAN PEMBAHASAN

4.1 Dataset

Data yang digunakan pada tugas akhir ini berupa data sekunder. Dataset diambil dari Electra dataset. Bagian ini akan menjelaskan fitur pada dataset, dan bagaimana penelitian ini memproses data agar dapat dimasukkan ke algoritma. Pertama-tama fitur. Electra Modbus dataset memiliki sebelas fitur. Tabel 4.1 menjelaskan fitur yang tersedia dalam dataset

Tabel 4.1
Fitur-Fitur dalam Dataset, Sumber: [11]

Fitur	Deskripsi	Tipe Data
Time	Waktu diambil data	String
Smac	Alamat MAC sumber	String
Dmac	Alamat MAC tujuan	String
Sip	Alamat IP sumber	String
Dip	Alamat IP tujuan	String
Request	Menunjukan apakah paket merupakan *request (master to slave)*	Boolean
fc	Function code	Integer
Error	Menunjukkan adanya error dalam pembacaan/penulisan	Boolean
Madd	*Memory Adress* untuk melakukan pembacaan/penulisan	Integer
data	Data yang dikirimkan oleh slave atau master	Integer
label	Menunjukan data normal dan data serangan	string

Semua fitur yang telah disebutkan tersedia dalam bentuk kolom yang dapat digunakan sebagai data latihan dan data untuk menguji algoritma yang telah disebutkan. Dalam fitur dataset ada fitur dengan nama label yang menunjukan apakah data bersifat normal atau serangan, agar dapat lebih rinci maka tabel 4.2 akan menunjukan proporsi data yang berada pada kolom label

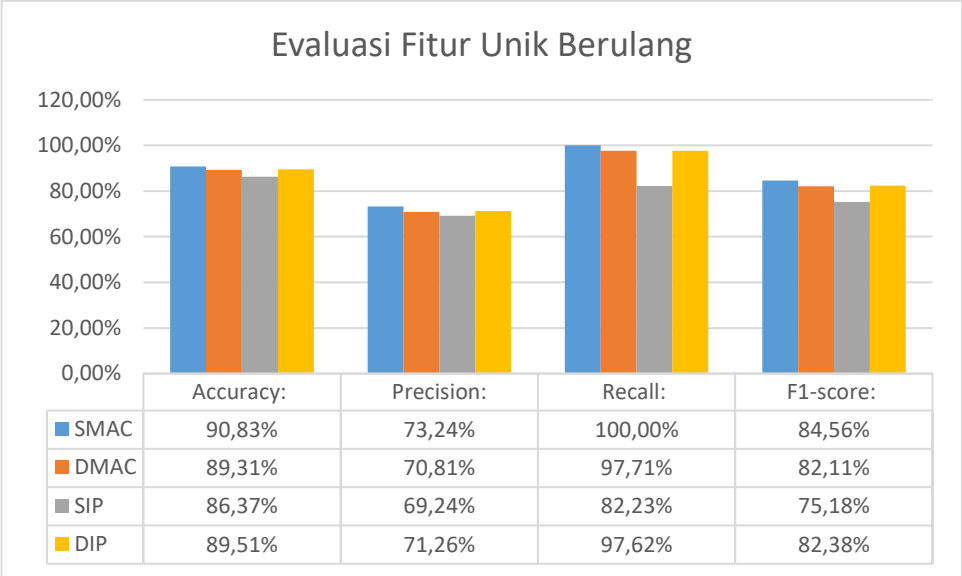
Tabel 4.2
Proporsi Kelas dalam Dataset, Sumber: [11]

Classes	Persentase sampel
Normal	94.8 %
Man-In-The-Middle Attack	9.52%
Function Code Recognition Attack	0.19%
Response Modification Attack	0.1%
Force error in response attack	0.007%
Read Attack	4.83%
Write Attack	0.06%
Replay Attack	0.006%

Agar dataset dapat dimasukkan ke dalam algoritma, diperlukan pre-processing terlebih dahulu. Dalam papernya Jehn-Ruey et.al menghilangkan data yang berulang. Hal ini karena dalam system kontrol industri proses kontrol dieksekusi ulang dari waktu ke waktu [12] oleh karena itu, duplikat dalam fitur Time akan dihilangkan. Selanjutkan akan dilakukan pre-processing agar data dapat diterima oleh LSTM, neural network tidak dapat menerima data lain selain angka yang artinya, data bersifat string akan dijadikan angka terlebih dahulu dengan *LabelEncoder* dari library Sklearn.PreProcessing. Pada metode validasi ID dataset, kolom selain kolom ID dan kolom label akan dihilangkan untuk dimasukkan ke dalam algoritma metode validasi. Setelah dilakukan pendeteksian maka kedua algoritma akan diuji performanya dengan membandingkan nilai hasil akurasi, presisi, recall, dan F1 Score. Untuk lebih jelasnya akan dijelaskan tiap kategori.

4.2 Hasil Metode Validasi ID

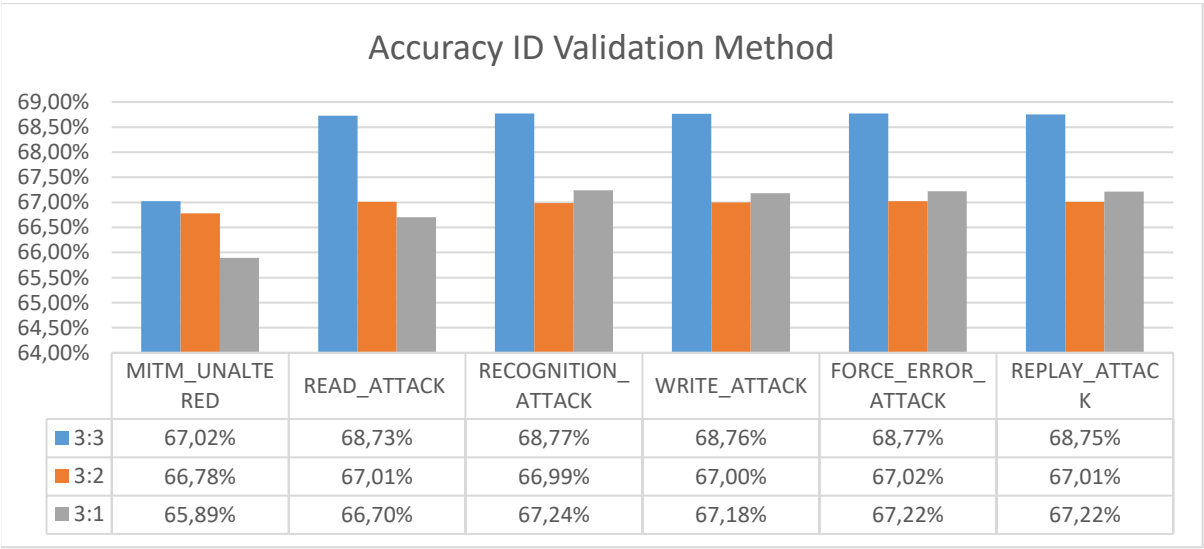
Agar metode validasi memiliki performa yang optimal, maka empat fitur dataset akan digunakan sebagai ID validasi. Keempat fitur ini adalah smac, dmac, sip, dan dip, keempat fitur ini digunakan karena alamat MAC unik secara global [13] dan alamat IP merupakan pengidentifikasi unik setiap perangkat dalam ModBus TCP [14]. Gambar 4.1 menjelaskan hasil dari evaluasi metode validasi ID untuk fitur yang dipilih.



	Accuracy:	Precision:	Recall:	F1-score:
SMAC	90,83%	73,24%	100,00%	84,56%
DMAC	89,31%	70,81%	97,71%	82,11%
SIP	86,37%	69,24%	82,23%	75,18%
DIP	89,51%	71,26%	97,62%	82,38%

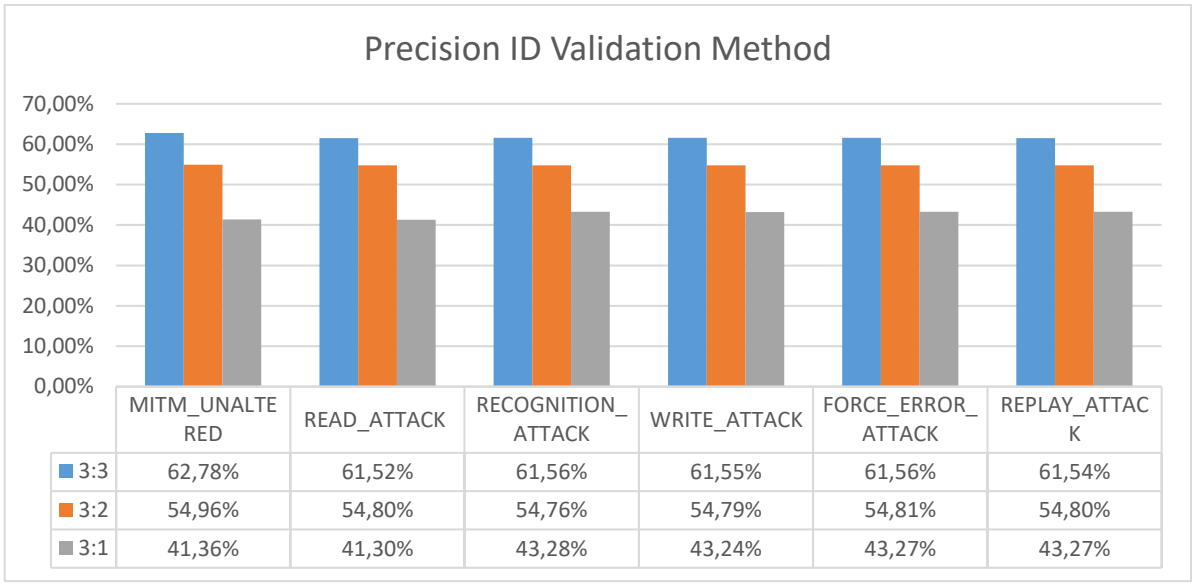
Gambar 4.1 Evaluasi Fitur Kandidat ID

Pada gambar 4.1 dapat dilihat bahwa fitur yang menghasilkan hasil terbaik adalah alamat smac artinya fitur smac cukup unik dan cukup banyak untuk pembuatan matriks transisi sah yang baik. Dengan terpilihnya smac sebagai fitur untuk ID, langkah selanjutnya untuk metode validasi adalah untuk mengetahui tipe serangan apa saja yang sulit dan mudah dideteksi oleh metode ini dengan cara mengisolasi tiap tipe serangan dengan data normal (contoh: NORMAL, MITM; NORMAL, REPLAY_ATTACK; dst). Telah disebutkan bahwa data kelas tidak seimbang sehingga dibuatlah rasio berbeda-beda untuk tiap serangan dengan data normal sebagai acuan utama. Rasio yang digunakan adalah 3:3, 3:2, 3:1 Berikut hasilnya



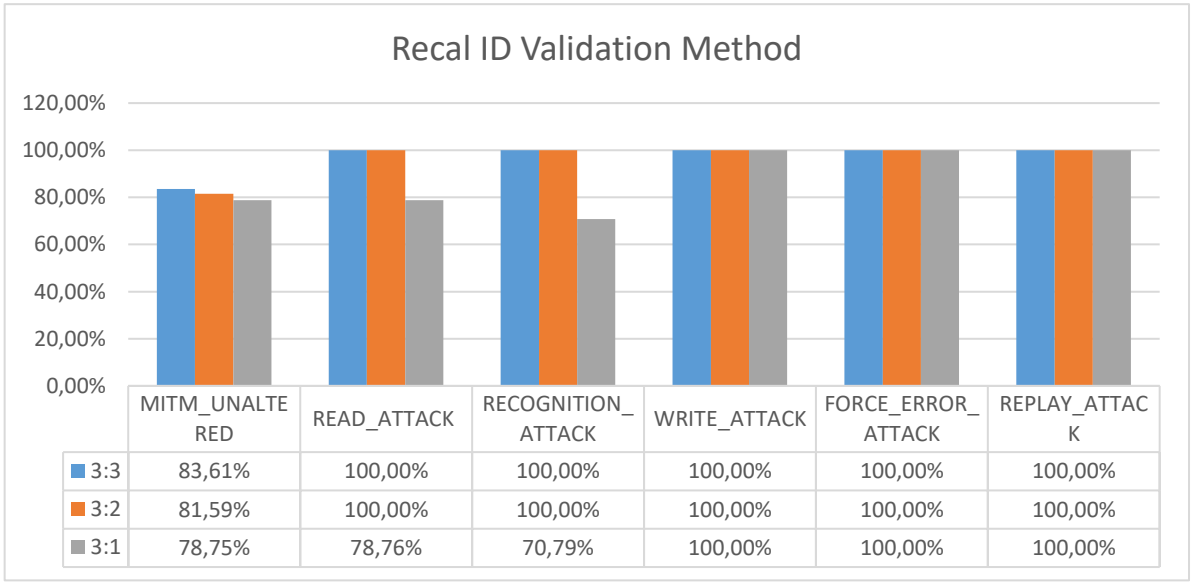
	MITM_UNALTERED	READ_ATTACK	RECOGNITION_ATTACK	WRITE_ATTACK	FORCE_ERROR_ATTACK	REPLAY_ATTACK
3:3	67,02%	68,73%	68,77%	68,76%	68,77%	68,75%
3:2	66,78%	67,01%	66,99%	67,00%	67,02%	67,01%
3:1	65,89%	66,70%	67,24%	67,18%	67,22%	67,22%

a. Evaluasi Akurasi



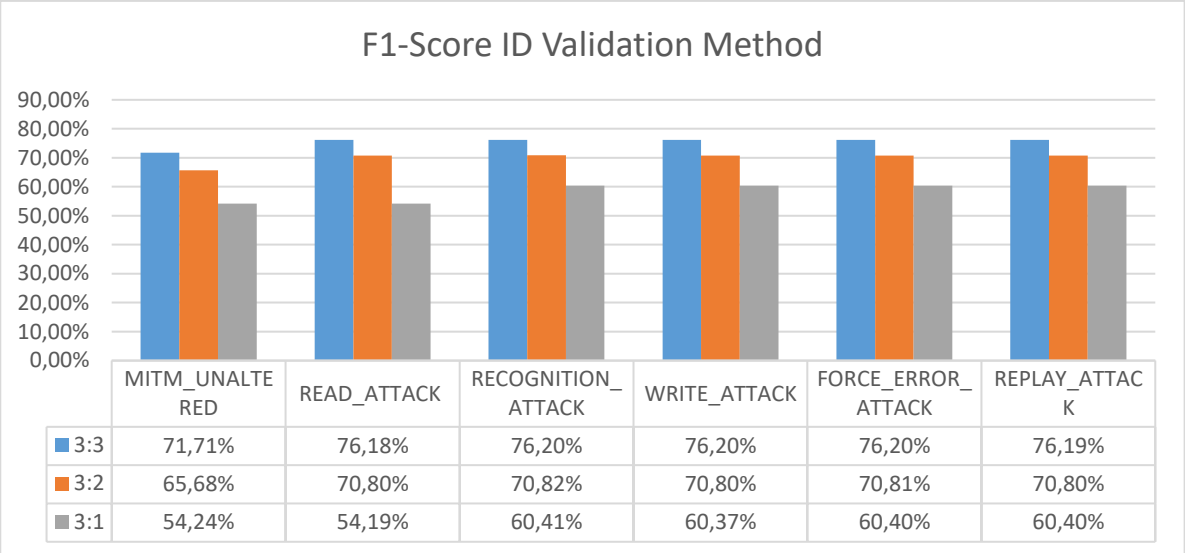
	MITM_UNALTERED	READ_ATTACK	RECOGNITION_ATTACK	WRITE_ATTACK	FORCE_ERROR_ATTACK	REPLAY_ATTACK
3:3	62,78%	61,52%	61,56%	61,55%	61,56%	61,54%
3:2	54,96%	54,80%	54,76%	54,79%	54,81%	54,80%
3:1	41,36%	41,30%	43,28%	43,24%	43,27%	43,27%

b. Evaluasi Presisi



	MITM_UNALTERED	READ_ATTACK	RECOGNITION_ATTACK	WRITE_ATTACK	FORCE_ERROR_ATTACK	REPLAY_ATTACK
3:3	83,61%	100,00%	100,00%	100,00%	100,00%	100,00%
3:2	81,59%	100,00%	100,00%	100,00%	100,00%	100,00%
3:1	78,75%	78,76%	70,79%	100,00%	100,00%	100,00%

c. Evaluasi Recall



	MITM_UNALTERED	READ_ATTACK	RECOGNITION_ATTACK	WRITE_ATTACK	FORCE_ERROR_ATTACK	REPLAY_ATTACK
3:3	71,71%	76,18%	76,20%	76,20%	76,20%	76,19%
3:2	65,68%	70,80%	70,82%	70,80%	70,81%	70,80%
3:1	54,24%	54,19%	60,41%	60,37%	60,40%	60,40%

d. Evaluasi F1-Score

Gambar 4.2 Evaluasi Metode Validasi ID dengan Rasio Data Serangan Berbeda. Dari a sampai dengan d adalah hasil evaluasi akurasi, presisi, recal dan skor F1

Dari hasil evaluasi tiap tipe serangan yang telah ditamplkan pada gambar 4.2, dapat dinyatakan bahwa tipe serangan yang sulit untuk dideteksi dengan metode validasi ID adalah MITM hal ini dapat dikarenakan alamat smac MITM meniru salah satu alamat smac normal menyebabkan *false positive* dalam validasi ID. Rasio Normal:Serangan berpengaruh terhadap hasil evaluasi metode validasi ID, dimana berkurangnya rasio serangan, semakin rendah juga hasil evaluasi metode ini.

4.3 Hasil Metode LSTM

Seperti sebelumnya metode perlu melakukan performa yang optimal salah satu cara menentukan model LSTM yang optimal adalah melakukan ablasi seperti pada penelitian yang dilakukan oleh Barbera, Cj et al. Dalam penelitian mereka, mereka menggunakan 20 dataset dan tiga metode neural network GRU, RNN, dan LSTM. Ablasi dilakukan pada ketiga neural network mereka dan didapatkan bahwa beberapa performa neural network meningkat dan ada beberapa yang tidak mengalami perubahan signifikan [29]. Pada penelitian ini, akan dilakukan ablasi layer LSTM dengan variasi layer 128, layer 64, dan layer 32. Akurasi akan dijadikan pembanding dalam ablasi layer LSTM dan akurasi terbaik akan menjadi pilihan utama untuk langkah selanjutnya.Tidak akan ada pemilihan fitur, karena LSTM merupakan neural network

yang dapat menggunakan input banyak, tetapi tetap akan dilakukan isolasi dan rasio antara data normal dan data serangan. Berikut adalah hasilnya.

Tabel 4.3
Hasil Evaluasi Akurasi LSTM 128

LSTM Layer	Attack Ratio			Attack name
	3:3	3:2	3:1	
128	99,63%	99,64%	99,64%	MITM_UNALTERED
	99,83%	99,80%	99,75%	READ_ATTACK
	99,83%	99,80%	99,75%	RECOGNITION_ATTACK
	99,83%	99,80%	99,75%	WRITE_ATTACK
	99,83%	99,80%	99,75%	FORCE_ERROR_ATTACK
	49,83%	59,80%	74,75%	REPLAY_ATTACK

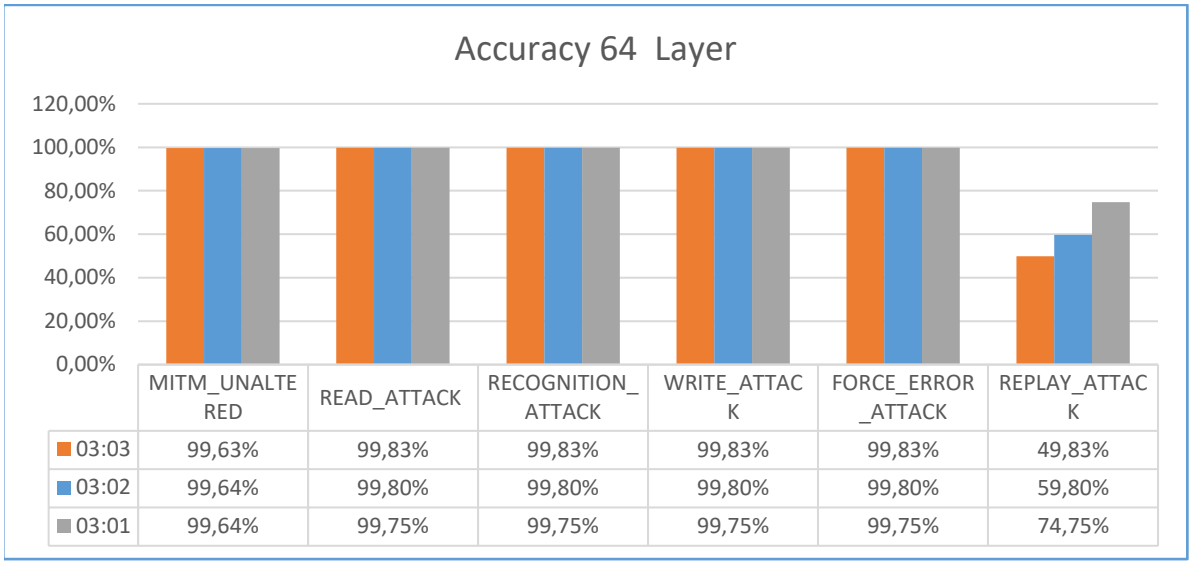
Tabel 4.4
Hasil Evaluasi Akurasi LSTM 64

LSTM Layer	Attack Ratio			Attack name
	3:3	3:2	3:1	
64	99,63%	99,64%	99,64%	MITM_UNALTERED
	99,83%	99,80%	99,75%	READ_ATTACK
	99,83%	99,80%	99,75%	RECOGNITION_ATTACK
	99,83%	99,80%	99,75%	WRITE_ATTACK
	99,83%	99,80%	99,75%	FORCE_ERROR_ATTACK
	49,83%	59,80%	74,75%	REPLAY_ATTACK

Tabel 4.5
Hasil Evaluasi LSTM 32

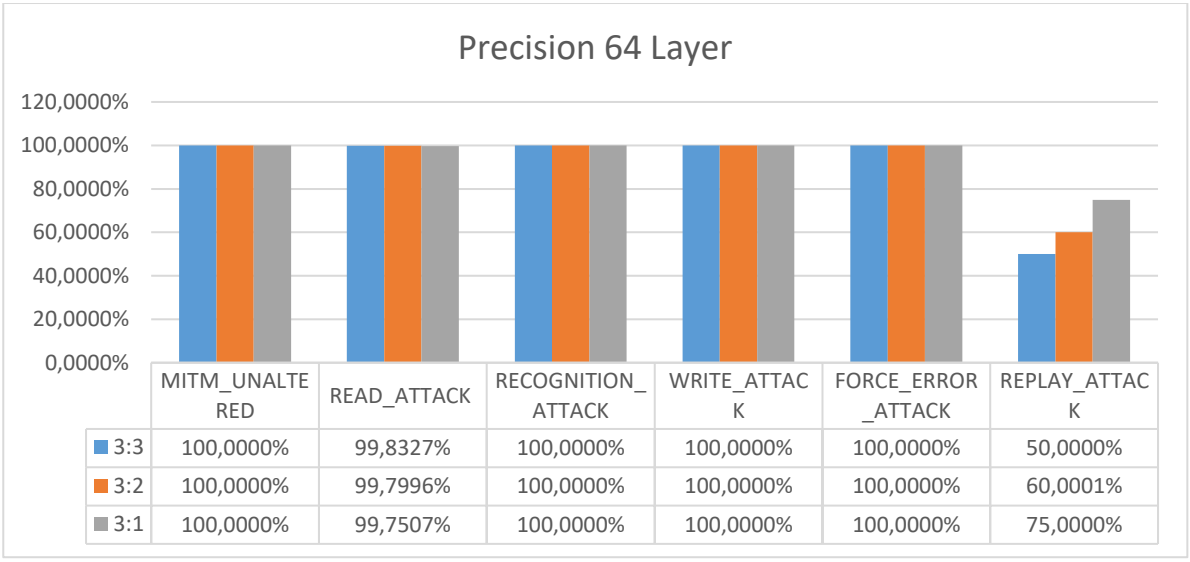
LSTM Layer	Attack Ratio			Attack name
	3:3	3:2	3:1	
32	99,63%	99,64%	99,64%	MITM_UNALTERED
	99,83%	99,80%	99,75%	READ_ATTACK
	99,83%	99,80%	99,75%	RECOGNITION_ATTACK
	99,83%	99,80%	99,75%	WRITE_ATTACK
	99,83%	99,80%	99,75%	FORCE_ERROR_ATTACK
	49,83%	59,80%	74,75%	REPLAY_ATTACK

Tabel 4.3 – Tabel 4.5 tidak menunjukkan perubahan performa setelah melakukan ablasi lapisan LSTM yang berarti ketiga lapisan LSTM valid dalam pengujian performa. Untuk tidak menghabiskan ruang dan mempersingkat waktu maka akan dipilih salah satu lapisan LSTM saja karena hasilnya akan sama. Lapisan LSTM yang akan dipilih adalah lapisan 64 karena lapisan 64 merupakan lapisan umum yang sering digunakan pada metode LSTM.



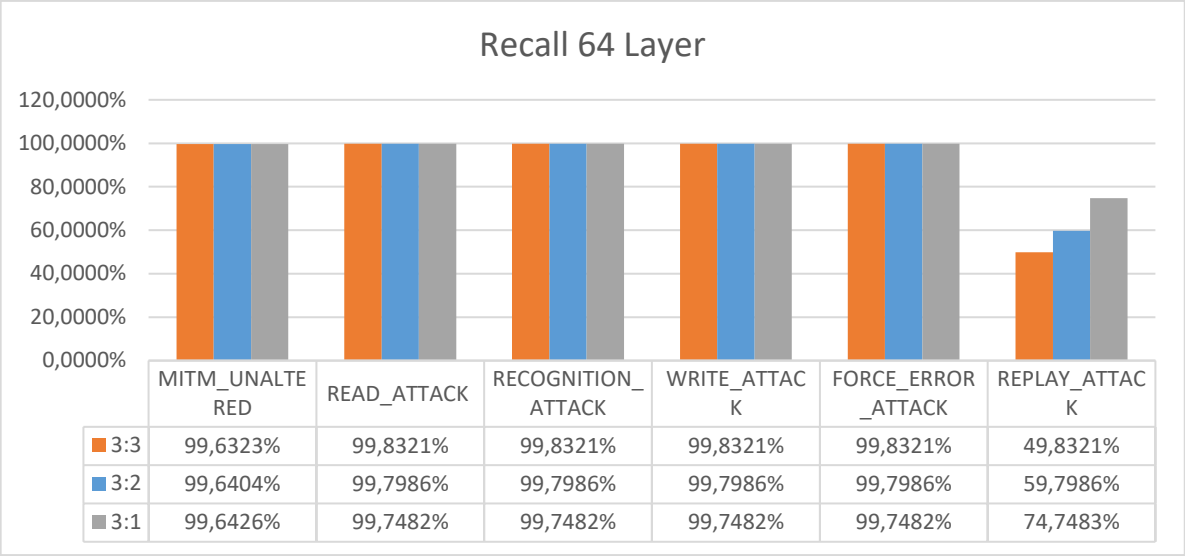
	MITM_UNALTERED	READ_ATTACK	RECOGNITION_ATTACK	WRITE_ATTACK	FORCE_ERROR_ATTACK	REPLAY_ATTACK
03:03	99,63%	99,83%	99,83%	99,83%	99,83%	49,83%
03:02	99,64%	99,80%	99,80%	99,80%	99,80%	59,80%
03:01	99,64%	99,75%	99,75%	99,75%	99,75%	74,75%

a. Hasil Akurasi LSTM



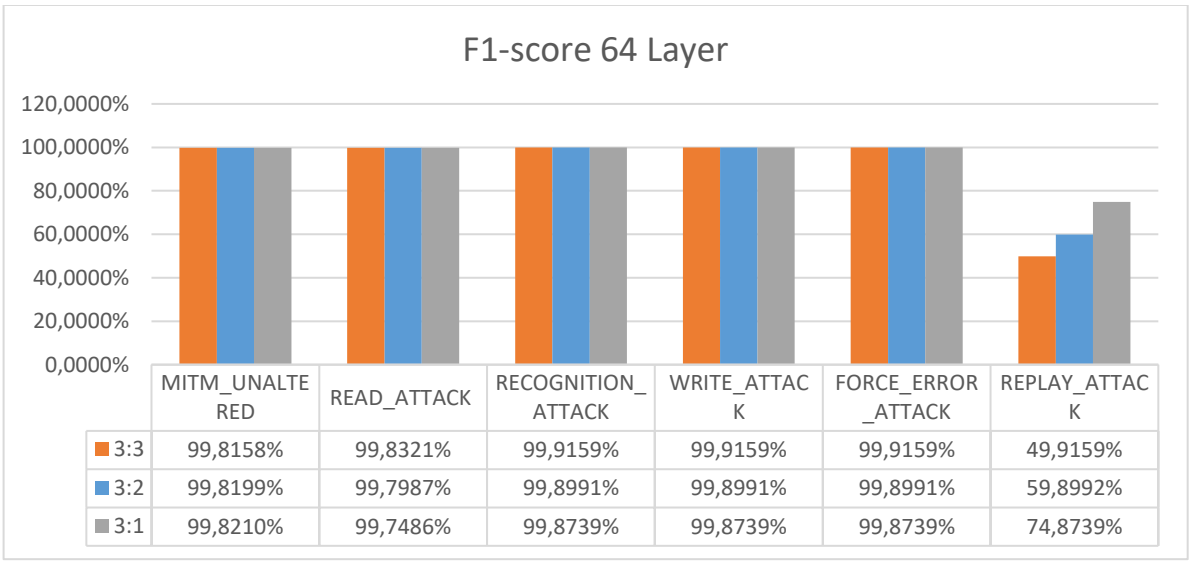
	MITM_UNALTERED	READ_ATTACK	RECOGNITION_ATTACK	WRITE_ATTACK	FORCE_ERROR_ATTACK	REPLAY_ATTACK
3:3	100,0000%	99,8327%	100,0000%	100,0000%	100,0000%	50,0000%
3:2	100,0000%	99,7996%	100,0000%	100,0000%	100,0000%	60,0001%
3:1	100,0000%	99,7507%	100,0000%	100,0000%	100,0000%	75,0000%

b. Hasil Presisi LSTM



	MITM_UNALTERED	READ_ATTACK	RECOGNITION_ATTACK	WRITE_ATTACK	FORCE_ERROR_ATTACK	REPLAY_ATTACK
3:3	99,6323%	99,8321%	99,8321%	99,8321%	99,8321%	49,8321%
3:2	99,6404%	99,7986%	99,7986%	99,7986%	99,7986%	59,7986%
3:1	99,6426%	99,7482%	99,7482%	99,7482%	99,7482%	74,7483%

c. Hasil Recal LSTM



	MITM_UNALTERED	READ_ATTACK	RECOGNITION_ATTACK	WRITE_ATTACK	FORCE_ERROR_ATTACK	REPLAY_ATTACK
3:3	99,8158%	99,8321%	99,9159%	99,9159%	99,9159%	49,9159%
3:2	99,8199%	99,7987%	99,8991%	99,8991%	99,8991%	59,8992%
3:1	99,8210%	99,7486%	99,8739%	99,8739%	99,8739%	74,8739%

d. Hasil F1-Score LSTM

Gambar 4.3 Hasil Evaluasi Performa LSTM

Evalusi LSTM menunjukkan bahwa LSTM susah mendeteksi *replay attack* hal ini dapat dikarenakan kurangnya pola yang berbeda. *Replay attack* melibatkan pengiriman ulang komunikasi sah yang direkam sebelumnya, membuat deteksi serangan ini termasuk sulit untuk algoritma LSTM yang mengandalkan sekuensi dalam data. Hal menarik lainnya adalah semakin rendah rasio data *replay attack* terhadap data normal, semakin tinggi juga hasil evaluasi performa LSTM.

4.4 Perbandingan Hasil Kedua Metode

Setelah melakukan evaluasi dengan parameter yang berbeda, terlihat bahwa hasil evaluasi metode LSTM jauh lebih bagus daripada hasil evaluasi metode validasi ID tetapi evaluasi sebelumnya dilakukan dengan mengisolasikan data normal dan data serangan. Perlu dilakukan evaluasi dengan semua data dalam dataset tanpa isolasi. Berikut adalah hasil evaluasi kedua metode

Tabel 4.6
Hasil Evaluasi Keseluruhan Kedua Metode

Evaluasi	Validasi ID	LSTM
Waktu	26s	6m 42s
Akurasi	90,83%	99,70%
Presisi	73,24%	99,70%
Recal	100,00%	99,70%
F1-score	87,75%	99,70%

Dari tabel 4.6 dapat dilihat bahwa LSTM menang jauh dari metode validasi sekuensi ID. Walaupu LSTM baik di evaluasi lainnya, validasi sekuensi ID termasuk cepat, sangat lebih cepat dari LSTM dan akurasinya, walaupun kalah, masih termasuk baik dengan hasil uji akurasi sebesar 90%. Masih perlu banyak pengempangan agar metode ini layak digunakan dalam protokol ModBus tetapi metode ini memiliki potensial.

Halaman ini sengaja dikosongkan

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Masih banyak kerentanan dalam protocol modbus, hal ini dikarenakan protokol tidak memasukkan otentikasi dan otorisasi, dan tidak ada verifikasi integritas data. Selain itu, semua data ditransfer dalam bentuk teks biasa, tanpa enkripsi apa pun.

Tujuan dari tugas akhir ini adalah mengevaluasi algoritma pendeteksi validasi sekuensi ID dan LSTM. Didapat kesimpulan sebagai berikut:

- Performa LSTM masih unggul dari metode validasi sekuensi ID dengan hasil, akurasi, presisi, recall, dan F1-skor sebesar 99,70%
- Metode validasi sekuensi ID memberikan hasil akurasi sebesar 90,83%, hasil presisi sebesar 73,24%, hasil recal sebesar 100,00%, dan hasil F1-Score sebesar 87,75%
- LSTM kesulitan mendeteksi serangan *replay attack*,
- Metode Validasi Sekuensi ID kesulitan mendeteksi serangan *man-in the-midde*,

5.2 Saran

Pada bagian ini akan dicantumkan saran kepada pembaca:

- Dataset walaupun memiliki banyak fitur kejanggalan masih terlalu sedikit. Disarankan mencari dataset yang memiliki fitur kejanggalan(serangan) yang lebih besar
- Pengembangan lebih lanjut terhadap metode matriks transisi dapat diakukan, seperti menambahkan variabel dalam matriks transisi sehingga mencapai akurasi yang lebih tinggi
- Diharapkan penelitian selanjutkan akan bisa mengambil data dari system modbus sendiri

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

[1] C. Parian, T. Guldimann and S. Bhatia, "Fooling the Master: Exploiting Weaknesses in the Modbus Protocol," *Third International Conference on Computing and Network Communications (CoCoNet'19),* pp. 2454-2458, 2020.

[2] J. Gao, L. Gan , F. Buschendorf, L. Zhang, H. Liu, P. Li, X. Dong and T. Lu, "LSTM for SCADA Intrusion Detection," *IEEE Pacific Rim Conference on Comunication*, 2019.

[3] J. Vávra and M. Hromada, "Comparison of the Intrusion Detection System Rules in," Springer International Publishing, [Online]. Available: https://core.ac.uk/download/pdf/43641216.pdf. [Accessed 11 Juli 2023].

[4] S. D. Anton, S. Kanoor, D. Fraunholz and H. D. Schotten, "Evaluation of machine learning-based anomaly detection algorithms on an industrial Modbus/TCP Data Set," 28 Mei 2019. [Online]. Available: https://arxiv.org/abs/1905.11757. [Accessed 11 Juli 2023].

[5] M. &. Marchetti and D. Stabili, "ANOMALY DETECTION OF CAN bus messages through analysis of ID sequences," *IEEE Intelligent Vehicles Symposium (IV)*, 2017.

[6] M. ÅKERMAN, Implementing Shop Floor IT for Industry 4.0, Chalmers University of Technology, 2018.

[7] K. Stouffer, J. Falco and K. Kent, Guide to Supervisory Control and Data Acquisition (SCADA) and Industrial Control Systems Security, National Institute, 2006.

[8] M. Al-Asiri and E.-S. M. El-Alfy, "On using physical based intrusion detection in SCADA systems," *Procedia Computer Science,* no. 170, pp. 34-42, 2020.

[9] S. Hong and M. Lee, "Challenges and direction toward secure communication in the SCADA system," *2010 8th Annual Communication Networks and Services Research Conference*, 2010.

[10] M. Inc., "“Modicon Modbus Protocol Reference Guide"," Juni 1996. [Online]. Available: https://modbus.org/docs/PI_MBUS_300.pdf. [Accessed 8 Juli 2023].

[11] B. Drury, Control Techniques Drives and Controls Handbook (2nd edition), Institution of Engineering and Technology, 2009.

[12] N. Erez and A. Wool, "Control variable classification, modeling and anomaly detection in Modbus/TCP SCADA systems," *International Journal of Critical Infrastructure Protection,* no. vol.10, pp. 59-70, 2015.

[13] W. L. Mostia, "Control," 4 Januari 2019. [Online]. Available: https://www.controlglobal.com/network/industrial-networks/article/11304326/introduction-to-modbus. [Accessed 8 July 2023].

[14] T. H. Morris and W. Gao, "Industrial Control System Cyber Attacks," *Electronic Workshops in Computing,* 2013.

[15] B. R. R. R. and P. A., "Intrusion Detection in SCADA Networks," *Lecture Notes in Computer Science,* p. 163–166, 2010.

[16] L. Fernández Maimó, A. Huertas Celdrán, Á. Perales Gómez, F. García Clemente, J. Weimer and I. Lee, "Intelligent and dynamic ransomware spread detection and mitigation in Integrated Clinical Environments," *Sensors,* p. 1114, 2019.

[17] L. Fernandez Maimo, A. L. Perales Gomez, F. J. Garcia Clemente, M. Gil Perez and G. Martinez Perez, "A self-adaptive deep learning-based system for anomaly detection in 5G networks," *IEEE Access,* pp. 7700-7712, 2018.

[18] M. Teixeira, T. Salman, M. Zolanvari, R. Jain, N. Meskin and M. Samaka, "SCADA system testbed for Cybersecurity Research Using Machine Learning Approach," *Future Internet,* p. 76, 2018.

[19] L. Rosa, T. Cruz, P. Simoes, E. Monteiro and L. Lev, "Attacking SCADA systems: A practical perspective," *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM),* 2017.

[20] Á. L. P. G. e. al, "On the Generation of Anomaly Detection Datasets in Industrial Control Systems," *IEEE Access,* pp. 177460-177473, 2019.

[21] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems,* pp. 2222-2232, 2017.

[22] R. Fu, Z. Zhang and L. Li, "Using LSTM and GRU Neural Network Method for Traffic Flow Prediction," *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC),* 2016.

[23] Á. L. PERALES GÓMEZ, L. F. MAIMÓ, A. H. CELDRÁN, F. . J. G. CLEMENTE, C. C. SARMIENTO, C. J. D. CANTO MASA and R. M. NISTAL, "On the Generation of Anomaly Detection," *IEEE Access*, 2019..

[24] S. Yan, "Medium," ML Review, 15 Nopember 2017. [Online]. Available: https://blog.mlreview.com/understanding-lstm-and-its-diagrams-37e2f46f1714. [Accessed 8 July 2023].

[25] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network," *THE EUROPEAN Physical Journal*, pp. 2313-2324, 2019.

[26] L. Salmela, N. Tsipinakis, A. Foi, C. Billet, G. Genty and J. M. Dudley, "Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network," *Nature Machine Intelligence*, vol. 4, no. 3, pp. 344-354, 2021.

[27] V. Bushaev, "Adam — latest trends in deep learning optimization.," Medium, 22 Oktober 2018. [Online]. Available: https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c. [Accessed 2023 July 23].

[28] J.-R. Jiang and Y.-T. Chen, "Industrial Control System Anomaly Detection and Classification Based on Network Traffic," *IEEE Access*, no. 10, pp. 41874-41888, 2022.

[29] J. Terra, "Simplilearn," 19 Juni 2023. [Online]. Available: https://www.simplilearn.com/what-is-mac-address-how-to-find-it-article#:~:text=The%20MAC%20is%20globally%20unique,a%206-byte%20hexadecimal%20number.. [Accessed 15 July 2023].

[30] "Modbus Gateways: connecting Modbus TCP and RTU Devices - EVALAN," 13 Februari 2023. [Online]. Available: https://evalan.com/modbus-gateways-connecting-modbus-tcp-and-rtu-devices/#:~:text=IP%20address%3A%20The%20IP%20address. [Accessed 16 Juli 2023].

[31] C. Barberan, S. Alemmohammad, N. Liu, R. Balestriero and R. Baraniuk, "NeuroView-RNN: It's about time," *2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022.

[32] Y. Gao, "Electronic Braking System of EV and HEV--Integration of Regenerative Braking, Automatic Braking Force Control and ABS," *42 Volt Technology and Advanced Vehicle Electrical Systems*, 2001.

[33] M. H. Westbrook, Development and future of battery, hybrid, and fuel-cell cars, London: The Institution of Electrical Engineers, 2005.

[34] R. Shiosansi, "Emissions Impacts and Benefits of Plug-in Hybrid Electric Vehicles and Vehicle-to-Grid Services," *Environmental Science Technology,* pp. 1199-1204, 2008.

[35] G. J. Offer, "Comparative analysis of battery electric, hydrogen fuel cell, and hybrid vehicle in a future sustainable road transport system," *energy policy,* vol. 38, pp. 24-29, 2010.

[36] O. Tur, "An Introduction to Regenerative Braking of Electric Vehicles as Anti-Lock Braking System," in *Proceedings of 2007 IEEE Intelligent Vehicles Symposium*, Istanbul, 2007.

[37] T. Murali, "Four Quadrant Operation and Control of Three Phase BLDC Motor," *International Conference on Circuits Power and Computing Technology,* 2017.

[38] A. Tashakori, "Modeling of BLDC Motor with Ideal Back-EMF for Automotive Applications," in *World Congress on Engineering*, London, 2011.

[39] C. P. K. S. Singh, "State-space Based Simulink Modeling of BLDC Motor and its Speed Control Using Fuzzy PID Controller," *International Journal of Advances in Engineering Science and Technology,* vol. 2, pp. 359-369, 2012.

[40] R. Errabelli, "Fault-Tolerant Voltage Source Inverter for Permanent Magnet Drives," *IEEE Transactions on Power Electronics,* vol. 27, 2012.

[41] K. Lubbers, “Design and Analysis of a Model Based Low Level Slip Controller Based on a Hybrid Braking System,” *Science in Systems and Control Delft University,* 2014.

[42] M. Blanke, Fault Tolerant Control Systems, London: Adventure Workd Press, 1999.

[43] L. Bai, "Electric Drive System with BLDC Motor," in *International Conference on Electric Information and Control Engineering*, Kuala Lumpur, 2011.

[44] K. Stouffer, S. Lightman, V. Pillitteri, M. Abrams and A. Hahn, 03 Juni 2015. [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-82/rev-2/final.

[45] M. S. D. G. A. &. C. M. Marchetti, ""Evaluation of anomaly detection for in-vehicle networks through information-theoretic algorithms"," *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leve,* (2016).

BIODATA PENULIS

I Gede Sanjaya Putra Vhyasa dilahirkan pada 30 Desember 2000 di Kota Banjarmasin, Kalimantan Selatan. Menempuh perkuliahan di Institut Teknlogi Sepuluh November, penulis berdomisili di Keputih Perintis. Merupakan anak pertama dari tiga bersaudara dari pasangan Bapak I Komang Dedy Suryanegara dan Ibu Berthy Yelly, penulis menempuh pendidikan Sekolah Dasar di SDN Sungai Miai 5 (2006-2012), lalu SMP Negeri 2 Banjarmasin (2012-2015), SMA Negeri 1 Banjarmasin (2015-2018), dan mulai menempuh pendidikan S1 Teknik Fisika FTIRS di ITS pada tahun 2019. Selama perkuliahan penulis aktif dalam organisasi UKM Kendo ITS dan sebagai asistem Laboratorium Sistem Tertanam dan Siber-Fisik. Pada bulan Juni 2023, penulis menyelesaikan tugas akhir dengan judul " **Evaluasi Algoritma Pendeteksi Penyusup Dengan Metode LSTM dan Validasi Sekuensi ID**". Jika ada kritik dan saran dapat menghubungi penulis melalui email: *sanjaya12bjm@gmail.com*. Atas kritik dan saran penulis mengucapkan terima kasih.