BAB 1
PENDAHULUAN

1.1 Latar Belakang

Dalam era digital yang semakin berkembang ini, jaringan dan infrastruktur teknologi informasi menjadi hal yang sangat penting bagi keberlangsungan operasional sebuah bisnis ataupun organisasi. Namun, seiring berjalannya waktu, sistem jaringan semakin lama akan semakin kompleks dan resiko ancaman siber menjadi meningkat, baik itu ancaman dari luar maupun dari dalam. Insiden seperti akses ilegal, kebocoran data, dan serangan siber lainnya sering disebabkan oleh akun yang sudah disusupi oleh pelaku penyerangan. Hal ini membuat serangan siber sulit di deteksi dengan metode tradisional seperti firewall atau antivirus karena pelaku menyamar dengan menggunakan akun karyawan tersebut. Sedangkan firewall atau antivirus ini hanya mencegah serangan dari luar dan serangan yang sudah diketahui saja. Untuk mencegah terjadinya serangan ini, solusinya adalah dengan menerapkan metode User and Entity Behavior Analytics (UEBA) (Al-mhiqani et al., n.d.).

User and Entity Behavior Analytics (UEBA) adalah sebuah pendekatan keamanan siber yang berfokus pada analisis perilaku pengguna (user) dan entitas (entity) secara terus-menerus. User adalah ‘manusianya’ dan entitas adalah ‘alatnya’. Sebagai contoh user adalah karyawan, administrator, atau bahkan pelanggan, sedangkan entity dapat berupa alamat IP, hostname, server, aplikasi, layanan, ataupun perangkat. Dengan memanfaatkan machine learning, UEBA membangun sebuah profil perilaku normal (baseline) untuk setiap entitas dalam jaringan. Ketika terjadi aktivitas yang menyimpang secara signifikan dari profil normal ini (outlier), sistem akan menandainya sebagai anomali atau potensi ancaman, bahkan jika aktivitas tersebut tidak melanggar aturan keamanan yang ada(Artioli et al., 2024).

Untuk merancang dan memvalidasi sistem UEBA ini, diperlukan sebuah lingkungan pengujian yang realistis dan berskala besar. Oleh karena itu, penelitian ini akan menggunakan dataset CESNET-timeseries2024. Dataset CESNET-timeseries2024 adalah sebuah dataset publik modern dari Jaringan Riset dan

Edukasi Nasional Ceko yang berisi data lalu lintas jaringan dunia nyata dari tahun 2024 (Koumar et al., 2025). Dataset ini awalnya masih mentah yang mencakup 66 miliar alur data (IP flows) dan 4 triliun paket data dengan total volume sekitar 3.7 Petabyte. Namun, setelah di agregasi berdasarkan waktu (contoh per 10 menit, per jam, dan per hari), data ini mencakup sekitar 275.000 alamat IP yang aktif dalam waktu kurun 40 minggu dengan volume 41.5 Gigabyte, dataset inilah yang akan digunakan pada penelitian ini.

Volume data yang sangat besar dari dataset ini menuntut penggunaan arsitektur Big Data untuk mendapatkan analisis yang lebih efisien (Haldorai et al., 2018). Arsitektur Big Data merujuk pada kerangka kerja atau susunan teknologi yang dirancang khusus untuk menangani tiga tantangan utama data dalam skala besar antara lain adalah Volume (ukuran sangat besar), Velocity (kecepatan data masuk yang tinggi), dan Variety (beragam jenis data). Maka dari itu, penelitian ini akan mengadopsi prinsip-prinsip arsitektur dari Apache Spot sebagai blueprint atau studi kasus.

Arsitektur Spot berbasis Apache Hadoop (sebagai penyimpanan), Apache Spark (sebagai mesin pemrosesan), dan algoritma machine learning untuk mendeteksi anomali yaitu Latent Dirichlet Allocation (LDA). Arsitektur ini menyediakan kerangka kerja yang fundamental dan transparan untuk diimplementasikan dalam penelitian ini (Towle, n.d.). Setelah dataset di evaluasi oleh sistem, hasilnya akan disajikan dalam bentuk dashboard visual menggunakan Apache Kibana agar dapat ditindaklanjuti oleh operator jaringan.

Apache Kibana adalah sebuah platform visualisasi data open-source yang dirancang untuk bekerja secara spesifik dengan datastore Elasticsearch. Elasticsearch itu sendiri adalah sebuah mesin pencari dan database analitik yang dioptimalkan untuk kecepatan dalam mengindeks, mencari, dan merangkum data dalam skala besar. Sebagai komponen visual utama dari Elastic Stack (ELK), fungsi utama Kibana adalah untuk mengubah data mentah berskala besar, seperti log jaringan dan data keamanan, menjadi wawasan yang mudah dipahami melalui berbagai macam grafik, tabel, peta, dan dashboard interaktif.

Dengan demikian, penelitian ini bertujuan untuk merancang, mengimplementasikan, dan mengevaluasi sebuah sistem UEBA untuk mendeteksi

anomali perilaku pada lalu lintas jaringan. Implementasi ini akan dilakukan dengan mereplikasi arsitektur analitik Apache Spot, mengujinya pada dataset CESNET-timeseries2024 untuk membuktikan efektivitasnya, serta menyajikan hasilnya dalam bentuk dashboard visual yang dapat ditindaklanjuti oleh operator jaringan menggunakan Apache Kibana.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, berikut merupakan permasalahan utama yang dihadapi dalam mengimplementasikan User and Entity Behavior Analytics (UEBA), antara lain:

1. Bagaimana cara mengimplementasikan sistem User and Entity Behavior Analytics (UEBA) menggunakan Apache Spot untuk mendeteksi anomali pada data lalu lintas jaringan?
2. Jenis - jenis perilaku anomali apa saja yang dapat diidentifikasi oleh sistem UEBA berbasis Apache Spot berdasarkan data yang dianalisis?
3. Bagaimana sistem UEBA tersebut membedakan karakteristik anomali dari pola perilaku normal yang telah dipelajari?
4. Seberapa efektif sistem UEBA berbasis Apache Spot yang dikembangkan dalam mendeteksi anomali menggunakan dataset CESNET-timeseries2024?

1.3 Batasan Masalah

Agar implementasi metode User and Entity Behavior Analytics (UEBA) dapat berjalan dengan baik dan benar, adapun batasan masalah yang harus ditetapkan, antara lain:

1. Penelitian ini fokus pada perancangan dan implementasi sistem UEBA dengan mereplikasi prinsip-prinsip arsitektur dari platform open-source Apache Spot sebagai studi kasus. Aspek konfigurasi dan alur kerja analitik akan dieksplorasi dalam konteks arsitektur tersebut.
2. Data yang digunakan untuk analisis dan pengujian sistem terbatas pada dataset publik CESNET-timeseries2024. Karakteristik dan jenis informasi dalam dataset ini akan menjadi dasar utama untuk deteksi anomali.
3. Deteksi anomali difokuskan pada perilaku yang dapat diidentifikasi dari fitur-fitur yang tersedia dalam dataset. Anomali yang memerlukan inspeksi konten

paket secara mendalam (deep packet inspection) berada di luar cakupan penelitian.

4. Penelitian ini akan memanfaatkan algoritma machine learning yang sudah ada dalam ekosistem pendukungnya (seperti Apache Spark) dan tidak bertujuan merancang algoritma baru dari dasar.
5. Implementasi dan pengujian sistem akan dilakukan dalam lingkungan laboratorium atau simulasi. Penelitian ini tidak melibatkan implementasi langsung pada jaringan produksi secara real-time.
6. Fokus utama sistem adalah pada deteksi anomali perilaku. Aspek respons insiden, investigasi forensik, atau mitigasi otomatis terhadap ancaman berada di luar cakupan penelitian ini.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Adapun tujuan dari penelitian ini sebagai berikut:

1. Mengimplementasikan sistem User and Entity Behavior Analytics (UEBA) menggunakan konsep platform Apache Spot dalam mendeteksi anomali pada lalu lintas jaringan.
2. Mengidentifikasi dan mengklasifikasikan jenis-jenis perilaku anomali yang dapat dideteksi oleh sistem UEBA berbasis Apache Spot dari dataset CESNET-timeseries2024.
3. Menganalisis dan menjelaskan bagaimana sistem UEBA yang dikembangkan tersebut mampu membedakan antara karakteristik perilaku anomali dengan pola perilaku normal pengguna dan entitas yang telah dipelajari dari data.
4. Mengevaluasi dan mengukur tingkat efektivitas serta performa sistem UEBA berbasis Apache Spot yang telah diimplementasikan dalam mendeteksi berbagai skenario anomali pada dataset CESNET-timeseries2024.

1.4.2 Manfaat

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan kontribusi terhadap literatur ilmiah di bidang keamanan siber, khususnya mengenai studi implementasi dan evaluasi sistem UEBA menggunakan konsep platform open-source seperti Apache Spot.
2. Menambah pemahaman mengenai kapabilitas dan tantangan dalam penerapan konsep Apache Spot untuk analisis perilaku dan deteksi anomali dalam skala data yang besar pada konteks keamanan jaringan.
3. Menyediakan analisis mengenai jenis-jenis anomali jaringan yang dapat teridentifikasi menggunakan pendekatan UEBA pada dataset spesifik (CESNET-timeseries2024), yang dapat berguna bagi peneliti lain.
4. Dapat menjadi dasar dan referensi bagi penelitian-penelitian selanjutnya yang ingin mengembangkan atau menguji metode serupa dalam deteksi ancaman siber.
5. Menghasilkan sebuah prototipe atau model kerja sistem UEBA menggunakan konsep platform Apache Spot yang dapat dijadikan panduan atau contoh bagi organisasi (khususnya skala kecil hingga menengah) yang ingin meningkatkan kapabilitas deteksi ancaman internal atau penyusupan akun dengan solusi yang relatif terjangkau.
6. Hasil analisis efektivitas sistem dapat memberikan wawasan kepada administrator jaringan dan praktisi keamanan mengenai potensi ancaman yang bisa dideteksi dan bagaimana UEBA dapat melengkapi sistem keamanan yang sudah ada.