BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dalam dunia kriptografi, penjagaan keamanan, keutuhan, dan integritas sebuah data menggunakan algoritma hash sangatlah dibutuhkan. Algoritma hash adalah sebuah metode kompres yang mengubah panjang data yang berbeda menjadi satu panjang yang ditetapkan [1]. Fungsi hash kriptografis juga digunakan sebagai mekanisme pemetaan data satu arah yang aman secara komputasional, sehingga mampu mempertahankan karakteristik keamanannya meskipun diterapkan pada data berukuran besar dengan karakteristik yang beragam [2]. Sebuah penelitian menyatakan bahwa beberapa algoritma hash sering digunakan untuk memverifikasi integritas sebuah data, mengonfirmasi file tersebut tidak pernah berubah [3].

Pada penelitian ini juga diperlukan sebuah data untuk di-*benchmark* seperti, variasi data, *Big Data*, dan pemakaian hash untuk sinkronisasi *cloud*. Variasi data seperti (gambar, suara, video, dan lain-lainnya) menunjukkan ledakan pertumbuhan yang sangat besar akhir–akhir ini [4]. Kemudian *Big Data* yang mengacu pada data dengan volume yang sangat besar, kecepatan pertumbuhan yang tinggi, keragaman jenis data, dan tingkat kompleksitas yang signifikan, sehingga tidak dapat dengan mudah diolah menggunakan metode atau perangkat lunak konvensional. Konsep ini juga mencakup penerapan teknik analisis data yang inovatif untuk mengungkap pola, tren, dan wawasan yang tersembunyi dalam data, sehingga informasi yang diperoleh dapat memberikan nilai tambah dan mendukung pengambilan keputusan secara lebih tepat [5].

Dalam praktik pengembangan sistem pada era ini, pemilihan fungsi hash kriptografis sering kali didasarkan pada standar dan tingkat adopsi yang luas, seperti penggunaan SHA-256, tanpa mempertimbangkan secara spesifik karakteristik data dan beban komputasi sistem. Padahal, pada lingkungan server dan komputasi awan yang memproses data dalam jumlah besar dan beragam, perbedaan karakteristik algoritma hash dapat mempengaruhi efisiensi penggunaan sumber daya seperti CPU, memori, dan waktu pemrosesan. Kondisi ini menunjukkan adanya kebutuhan untuk mengevaluasi kinerja dan stabilitas berbagai fungsi hash terhadap variasi data dan beban sistem secara lebih sistematis. Sebuah penelitian menyatakan salah satu algoritma yang umum

dipakai untuk hashing adalah *SHA-256* yang merupakan salah satu anggota *Secure Hash Algorithm (SHA)* dimana algoritma ini mampu menghasilkan nilai hash yang unik [6]. Fungsi hash kriptografis seperti *SHA-256* dan telah distandardisasi dalam *Federal Information Processing Standard (FIPS)* 180-4 oleh *NIST*, dan karena status standarnya ini, algoritma tersebut banyak diterapkan dalam berbagai skenario keamanan informasi serta sistem nyata sebagai mekanisme untuk menghasilkan digest yang tahan terhadap perubahan data [7].

Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi stabilitas berbagai fungsi hash kriptografis terhadap variasi karakteristik data input. Evaluasi dilakukan melalui dua dimensi utama, yaitu *Resource Stability* dan *Statistical Stability*. *Resource Stability* digunakan untuk mengukur bagaimana kinerja algoritma hash dalam memanfaatkan sumber daya komputasi, yang direpresentasikan melalui metrik penggunaan CPU, Memory, Latency, dan Throughput. Dimensi ini menunjukkan sejauh mana suatu algoritma mampu mempertahankan performa yang konsisten ketika memproses data dengan karakteristik yang berbeda-beda. Pendekatan serupa telah digunakan dalam literatur untuk menilai efisiensi komputasi fungsi hash, misalnya dalam perbandingan efisiensi *SHA-256*, *SHA-3*, dan *BLAKE2* [8]. Sementara itu, *Statistical Stability* digunakan untuk mengevaluasi kualitas dan konsistensi sifat kriptografis dari output hash, yang meliputi *Avalanche Effect*, *Collision Rate*, distribusi output, serta bias statistik. Dimensi ini menggambarkan seberapa stabil dan andal suatu algoritma hash dalam menjaga sifat acak dan ketahanannya terhadap variasi input. Evaluasi empiris terhadap *avalanche effect* dan *Bit Independence Criterion* pada berbagai fungsi hash telah dilakukan dalam literatur untuk menilai randomisasi keluaran dan resistansi terhadap pola statistik tertentu [9]. Dengan mengombinasikan kedua dimensi tersebut, penelitian ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai stabilitas masing-masing algoritma hash, sehingga pemilihan algoritma dapat dilakukan secara lebih tepat sesuai dengan karakteristik data dan kebutuhan sistem.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Apa pengaruh variasi karakteristik data input (seperti teks, gambar, audio, dan video) terhadap kinerja dan sifat statistik dari berbagai algoritma hash kriptografis?
2. Bagaimana tingkat *Resource Stability* dari masing-masing algoritma hash kriptografis ketika memproses data dengan ukuran dan karakteristik yang berbeda, yang diukur melalui penggunaan CPU, Memory , Latency, dan Throughput?
3. Bagaimana tingkat *Statistical Stability* dari masing-masing algoritma hash kriptografis terhadap variasi data input, ditinjau dari *Avalanche effect*, *Collision rate*, distribusi output, dan bias statistik?
4. Apakah terdapat perbedaan signifikan dalam stabilitas antara algoritma hash yang telah distandarisasi (seperti *SHA-256*) dan algoritma hash lain yang memiliki desain dan karakteristik internal berbeda?
5. Algoritma hash manakah yang menunjukkan stabilitas paling konsisten ketika diterapkan pada berbagai jenis dan ukuran data dalam lingkungan komputasi yang menyerupai sistem server atau komputasi awan?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Menganalisis stabilitas kinerja (*Resource Stability*) dari berbagai algoritma hash kriptografis dalam memproses data dengan ukuran dan karakteristik yang berbeda-beda, yang diukur melalui penggunaan CPU, *Memory*, *Latency*, dan *Throughput*.
2. Menganalisis stabilitas statistik (*Statistical Stability*) dari algoritma hash kriptografis terhadap variasi data input, berdasarkan A*valanche effect*, *Collision rate*, Distribusi output *(hexadecimal)*, dan Bias statistik *(Bit)*.
3. Membandingkan algoritma hash yang telah distandarisasi (seperti *SHA-256*) dengan algoritma hash lain yang memiliki desain internal berbeda dalam hal stabilitas sumber daya dan stabilitas statistik.

4. Mengidentifikasi algoritma hash yang paling konsisten dan stabil ketika digunakan pada berbagai jenis data (teks, gambar, musik, dan video) dalam lingkungan komputasi yang menyerupai sistem server.
5. Menyediakan dasar evaluasi teknis yang dapat digunakan untuk memilih algoritma hash yang paling sesuai untuk sistem penyimpanan data, verifikasi integritas, dan sistem berbasis *cloud*.

1.3.2 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi terhadap kajian kriptografi terapan, khususnya dalam memahami hubungan antara karakteristik data dan stabilitas algoritma hash.
2. Menyediakan kerangka analisis baru berupa konsep *Resource Stability* dan *Statistical Stability* sebagai pendekatan kuantitatif dalam mengevaluasi algoritma hash.
3. Memperkaya literatur penelitian tentang evaluasi performa fungsi hash yang tidak hanya berfokus pada aspek keamanan, tetapi juga pada stabilitas dan efisiensi sistem.
4. Menjadi referensi teknis bagi perusahaan teknologi, layanan hosting, dan penyedia *cloud* dalam merancang sistem penyimpanan dan verifikasi data yang lebih andal.
5. Memberikan dasar pengambilan keputusan dalam penggunaan algoritma hash pada sistem dengan beban tinggi dan variasi data yang besar, seperti sistem arsip digital, *blockchain*, dan *big data*.

1.4 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya membahas fungsi hash kriptografis dan tidak mencakup algoritma enkripsi, tanda tangan digital, atau mekanisme kriptografi lainnya.
2. Algoritma hash yang dianalisis dibatasi pada beberapa algoritma yang umum digunakan dan representatif seperti *SHA-256, BLAKE2, Whirlpool,* dan *Skein* sehingga tidak mencakup seluruh algoritma hash yang ada.
3. Data uji dalam penelitian ini dibatasi pada empat jenis data digital, yaitu teks (±2 GB), musik (±7 GB), video (±7 GB), dan gambar (±8 GB). Pembatasan ini dilakukan untuk mengontrol kompleksitas pengujian dan memastikan proses

eksperimen dapat dilakukan secara terukur, sehingga hasil penelitian tidak merepresentasikan seluruh variasi data digital di dunia nyata.

4. Evaluasi *Resource Stability* dibatasi pada metrik *CPU Usage, Memory Usage, Latency, dan Throughput*, tanpa memasukkan faktor lain seperti konsumsi energi atau suhu perangkat.
5. Evaluasi *Statistical Stability* dibatasi pada pengukuran *Avalanche effect, Collision rate,* distribusi output *(Hexadecimal),* dan bias statistik *(Bit)* tanpa membahas secara mendalam aspek kriptanalisis tingkat lanjut.

1.5 Keterbaruan

Penelitian ini memiliki beberapa aspek keterbaruan yang membedakannya dari penelitian sebelumnya, yaitu sebagai berikut:

1. Penelitian ini menggunakan dataset buatan sendiri berupa file acak dengan ukuran bervariasi untuk menguji performa dan stabilitas algoritma hash. Pendekatan ini memungkinkan kontrol penuh terhadap karakteristik data input, sehingga evaluasi terhadap *Resource Stability* dan *Statistical Stability* dapat dilakukan secara sistematis tanpa tergantung pada data publik yang mungkin memiliki bias tertentu.
2. Salah satu keterbaruan dalam penelitian ini juga ditunjukkan melalui pengujian algoritma *SHA-256* pada empat jenis data yang berbeda secara simultan, yaitu foto, audio, video, dan teks, sehingga mampu memberikan wawasan baru terkait stabilitas nilai hash dan kinerja algoritma pada berbagai format data yang masih jarang diuji secara menyeluruh, dimana pada penelitian sebelumnya hanya menguji terhadap file audio saja [10].
3. Dalam penelitian ini, salah satu dataset yang digunakan adalah *FMA (Free Music Archive)* yang tersedia di *GitHub*, berukuran sekitar 7 GB. Dataset ini berisi koleksi musik berlisensi bebas yang mencakup berbagai genre dan durasi, sehingga cocok untuk menguji algoritma hash terhadap variasi data audio berskala besar. Dataset ini juga merepresentasikan kondisi pemrosesan data multimedia pada lingkungan server atau *cloud*, memberikan konteks realistis dalam pengukuran stabilitas algoritma hash [11].