UNIVERSITAS
GADJAH MADA
Rancang Bangun Pemindai Hardcoded Secrets Berbasis Shannon Entropy dan Random Forest untuk Mengurangi False Positive
Heaven Aulianisa Pambudi Putri, Dr. Ir. Ronald Adrian S.T., M.Eng., IPM.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

"Rancang Bangun Pemindai *Hardcoded Secrets* Berbasis Shannon Entropy dan Random Forest untuk Mengurangi *False Positive*"

Heaven Aulianisa Pambudi Putri
22/492474/SV/20511

Penyisipan kredensial rahasia secara langsung pada kode sumber atau *harcoded secrets* merupakan celah keamanan yang berisiko tinggi dalam pengembangan perangkat lunak. Kredensial seperti API Key, token, dan *password* dapat terekspos apabila tersimpan di repositori. Pemindai keamanan konvensional berbasis ekspresi reguler (*Regular Expression*/*regex*), yang memiliki kelemahan fundamental berupa tingginya tingkat *False Positive* akibat ketidakmampuan membedakan kredensial asli dengan teks biasa, *placeholder*, atau *hash* statis. Kondisi ini pada akhirnya memicu *alert fatigue* pada analis keamanan. Penelitian ini bertujuan merancang dan membangun Secret Finder, yaitu sistem pemindai *harcoded secrets* yang mengombinasikan *regex*, Shannon Entropy, dan algoritma Random Forest untuk mereduksi *false positive,* serta mengklasifikasikan tingkat keparahan temuan ke dalam kategori *High*, *Medium*, dan *Low*. Metode yang digunakan terdiri atas pencocokan *regex* sebagai penyaring awal, perhitungan Shannon Entropy untuk mengukur tingkat keacakan *string*, serta Random Forest Classifier untuk menentukan tingkat bahaya berdasarkan fitur hasil ekstraksi. Hasil pengujian menunjukkan bahwa pemindaian berbasis *regex* menghasilkan *false positive rate* sebesar 77,78%, sedangkan integrasi Shannon Entropy dan Random Forest mampu menurunkan *false positive* sebesar 91,07% pada mode *Medium*. Sistem pada mode *High* menghasilkan *false positive rate* sebesar 0% dalam skenario pengujian dan mampu mereduksi beban pemeriksaan manual sebesar 90,28% dari 72 temuan mentah menjadi 7 peringatan prioritas tinggi.

Kata kunci: *Harcoded Secrets, False Positive,* Shannon Entropy, Random Forest, *Reguler Expression.*

UNIVERSITAS
GADJAH MADA
Rancang Bangun Pemindai Hardcoded Secrets Berbasis Shannon Entropy dan Random Forest untuk Mengurangi False Positive
Heaven Aulianisa Pambudi Putri, Dr. Ir. Ronald Adrian S.T., M.Eng., IPM.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

"Design and Development of a Hardcoded Secrets Scanner Based on Shannon Entropy and Random Forest to Reduce False Positives"

Heaven Aulianisa Pambudi Putri
22/492474/SV/20511

Embedding sensitive credentials directly into source code or hardcoding secrets is a high-risk security vulnerability in software development. Credentials such as API keys, tokens, and passwords can be exposed if stored in a repository. Conventional security scanners based on regular expressions (regex) have a fundamental weakness: a high rate of false positives due to their inability to distinguish genuine credentials from plain text, placeholders, or static hashes. This ultimately leads to alert fatigue among security analysts. This research aims to design and build Secret Finder, a system for scanning hardcoded secrets that combines regex, Shannon Entropy, and the Random Forest algorithm to reduce false positives and classify the severity of findings into High, Medium, and Low categories. The methods used consist of regex matching as an initial filter, Shannon Entropy calculation to measure the string's randomness level, and a Random Forest Classifier to determine the threat level based on extracted features. Test results show that regex-based scanning yields a false positive rate of 77.78%, while the integration of Shannon Entropy and Random Forest reduces false positives by 91.07% in Medium mode. In High mode, the system produced a false positive rate of 0% in the test scenario and was able to reduce the manual inspection workload by 90.28%, from 72 raw findings to 7 high-priority alerts.

Keyword: Harcoded Secrets, False Positive, Shannon Entropy, Random Forest*, Reguler Expression.*