ABSTRAK

Blockchain adalah sistem pencatatan data digital yang terdesentralisasi di banyak komputer sehingga setiap transaksi bersifat transparan, permanen, dan sulit dimanipulasi tanpa persetujuan jaringan. Di atas blockchain seperti Ethereum berjalan *smart contract*, yaitu program yang tersimpan dan dieksekusi otomatis oleh jaringan untuk menjalankan kesepakatan tanpa perantara; karena mengelola aset bernilai finansial, smart contract menjadi sasaran sekaligus alat bagi pihak jahat. Mayoritas *smart contract* yang ter-deploy pada *blockchain* Ethereum tidak mempublikasikan kode sumbernya, sementara kontrak *malicious* terus menimbulkan kerugian finansial yang besar; akibatnya, deteksi kontrak berbahaya harus dapat dilakukan langsung pada level *bytecode*. Penelitian ini mengimplementasikan metode deteksi *malicious smart contract* berbasis *opcode* menggunakan Graph Neural Network (GNN) dengan representasi Control Flow Graph (CFG) pada *dataset* Forta Network yang terdiri dari 139.600 kontrak dengan rasio ketidakseimbangan kelas 936:1 (149 *malicious* berbanding 139.451 *benign*). Setiap *smart contract* direpresentasikan sebagai CFG di mana *basic block* menjadi *node* berfitur 22 dimensi yang terdiri dari frekuensi 14 sensitive *opcode* dan representasi *one-hot* tipe instruksi exit, sedangkan hubungan antar blok direpresentasikan sebagai *edge* dengan lima kategori. Graf yang terbentuk diproses menggunakan arsitektur GCNConv tiga lapis dengan TopK *Pooling* dan *readout* mean+max. Untuk menangani ketidakseimbangan kelas, diterapkan *oversampling* berbasis Gaussian *noise* dengan rasio 50 dan pembobotan kelas pada fungsi CrossEntropyLoss. Sebagai pembanding, metode klasifikasi tradisional berbasis frekuensi *opcode* direplikasi pada *dataset* yang sama menggunakan kondisi evaluasi yang identik. Pada kondisi evaluasi setara dengan *test set* 56 *malicious* dan 1.204 *benign*, metode GNN mencapai F1-score 0,8544, Precision 0,9362, Recall 0,7857, dan AUC 0,9275. Metode klasifikasi tradisional terbaik (KNN) mencapai F1-score 0,8704 dengan Precision 0,9038. Metode GNN unggul dalam Precision sehingga menghasilkan lebih sedikit false positive, sementara KNN unggul dalam Recall. Hasil penelitian menunjukkan bahwa representasi berbasis CFG yang diproses GNN mampu menangkap pola alur eksekusi *bytecode* yang tidak dapat direpresentasikan oleh pendekatan frekuensi *opcode* secara langsung.

Kata kunci: *smart contract malicious*, Control Flow Graph, Graph Neural Network, GCNConv, deteksi *bytecode*, Ethereum, ketidakseimbangan kelas

ABSTRACT

A blockchain is a decentralized digital record-keeping system distributed across many computers, making every transaction transparent, permanent, and difficult to tamper with without network consensus. On blockchains such as Ethereum run smart contracts—programs stored on and automatically executed by the network to enforce agreements without intermediaries; because they manage financially valuable assets, smart contracts are both a target and a tool for malicious actors. Most smart contracts deployed on the Ethereum blockchain do not publish their source code, while malicious contracts continue to cause substantial financial losses; consequently, detection must operate directly at the bytecode level. This study implements an opcode-based malicious smart contract detection method using a Graph Neural Network (GNN) with a Control Flow Graph (CFG) representation on the Forta Network dataset, which consists of 139,600 contracts with a class imbalance ratio of 936:1 (149 malicious and 139,451 benign). Each smart contract is represented as a CFG where basic blocks become nodes with 22dimensional features comprising the frequency of 14 sensitive opcodes and a onehot encoding of exit instruction types, while control flow between blocks is encoded as edges with five categories. The resulting graphs are processed using a three-layer GCNConv architecture with TopK Pooling and mean+max readout. Class imbalance is addressed through Gaussian noise-based oversampling at ratio 50 and class weighting in the CrossEntropyLoss function. For comparison, a traditional opcode frequency-based classification method is replicated on the same dataset under identical evaluation conditions. Under equivalent (apple-to-apple) evaluation conditions on a test set of 56 malicious and 1,204 benign contracts, the GNN method achieves an F1-score of 0.8544, Precision of 0.9362, Recall of 0.7857, and AUC of 0.9275. The best traditional classifier (KNN) achieves an F1-score of 0.8704 with Precision of 0.9038. The GNN method produces fewer false positives, while KNN yields higher Recall. Results indicate that CFG-based representations processed by GNN can capture bytecode execution flow patterns that opcode frequency approaches cannot directly represent.

Keywords: malicious smart contract, Control Flow Graph, Graph Neural Network, GCNConv, bytecode detection, Ethereum, class imbalance