Secure Smart Contracts: Prevention & Detection with the aid of Deep Learning

Abstract

Blockchain technologies have revolutionized many areas across finance and computer science, by offering a network that enables decentralized solutions as well as ensures security and accountability. With the increase in the adoption of blockchain technology in providing decentralized solutions to various problems, smart contracts have become incredibly popular - to the point that billions of US Dollars are currently exchanged every day through such technology. Meanwhile, various vulnerabilities in smart contracts have been exploited by attackers to steal cryptocurrencies worth millions of dollars. In this tutorial we will explore state-of-the-art techniques for both prevention and detection in the context of securing smart contracts.

The automatic detection of smart contract vulnerabilities is an essential research problem. Existing solutions to this problem particularly rely on human experts to define features or different rules to detect vulnerabilities, which can be very error-prone. In this tutorial, we explore recent work towards automating this task in an effective manner through machine learning.

For preventing vulnerabilities entirely, methods such as Ethereum bytecode rewriting and validation method will be presented and evaluated for securing smart contracts in decentralized cryptocurrency systems without accessing the contract source code.

Intended audience

The intended audience for this can range from beginners who will learn some important concepts in blockchains and smart contracts, to deep learning practitioners who wish to work in the field of smart contracts and blockchain security and apply deep learning techniques in this domain.

Tutorial Speakers

1. **Dr. Latifur Khan**

Dr. Latifur Khan is currently a full Professor (tenured) in the Computer Science department at the University of Texas at Dallas, USA where he has been teaching and conducting research since September 2000.
He received his Ph.D. degree in Computer Science from the University of Southern California (USC) in August of 2000.
Dr. Khan is an ACM Distinguished Scientist and received IEEE Big Data Security Senior Research Award, in May 2019, and Fellow of SIRI (Society of Information Reuse and Integration) award in Aug, 2018. He has received prestigious awards including the IEEE Technical Achievement Award

for Intelligence and Security Informatics, IEEE Big Data Security Award and IBM Faculty Award (research) 2016.

Dr. Latifur Khan has published over 300 papers in premier journals such as VLDB, Journal of Web Semantics, IEEE TDKE, IEEE TDSC, IEEE TSMC, and AI Research and in prestigious conferences such as AAAI, IJCAI, ACM WWW, CIKM, ICDE, ACM GIS, IEEE ICDM, IEEE BigData, ECML/PKDD, PAKDD, ACM Multimedia, ICWC, ACM SACMAT, IEEE ICSC, IEEE Cloud and INFOCOM. He has been invited to give keynotes and invited talks at a number of conferences hosted by IEEE and ACM. In addition, he has conducted tutorial sessions in prominent conferences such as SIGKDD 2017, 2016, IJCAI 2017, AAAI 2017, SDM 2017, PAKDD 2011 & 2012, DASFAA 2012, ACM WWW 2005, MIS2005, and DASFAA 2007.

Currently, Dr. Khan's research area focuses on big data management and analytics, data mining and its application over cyber security, complex data management including geo-spatial data and multimedia data. His research has been supported by grants from NSF, NIH, the Air Force Office of Scientific Research (AFOSR), DOE, NSA, IBM and HPE. More details can be found at: www.utdallas.edu/~lkhan/

2. **Feng Mi**

Feng Mi is a PhD student at the University of Texas at Dallas. He previously completed a Master of Science in Computer Engineering at the University of Delaware, as well as a Bachelor of Engineering at the Nanjing University of Information Science and Technology. His research lies broadly in machine learning, including areas such as metric learning, as well as cybersecurity.

3. **Sadaf MD Halim**

Sadaf MD Halim is a PhD student at the University of Texas at Dallas. He also previously completed his Master of Science in Computer Science at UT Dallas, and he obtained a Bachelor of Science in CS at the Islamic University of Technology (IUT) in Bangladesh. His research interests include blockchains, federated learning, and generally learning in a distributed setting.

Technical Issues Addressed, and their timeliness

Smart contracts are responsible for large-volume transactions in blockchains, including Ethereum. These contracts are an essential part of automating a lot of procedures that are relevant to transactions, which is why they have enjoyed widespread popularity. This means that ensuring their security is of paramount importance. Manually verifying and checking for vulnerabilities in a smart contract is a tedious process and the content of this tutorial shows how it can be automated with the aid of deep learning and some creative techniques. This tutorial will also present methods to prevent vulnerabilities through methods such as bytecode rewriting and validation.

Outline and Schedule

This tutorial will begin with a few blockchain basics, before moving into the Ethereum Blockchain and smart contracts. Next, it will explore typical vulnerabilities experienced by smart contracts. This will motivate the core content of the tutorial - a framework that analyzes and detects vulnerabilities in the Ethereum smart contract platform. We discuss each section in brief below:

- Blockchain basics, smart contracts (40 minutes): The first part of the tutorial will be a brief introduction to blockchain basics and its operations. Brief explanations of the basics of blockchain such as proof of work will be provided. We will explain what blockchains can and cannot achieve, and then explain what motivated smart contracts. We will briefly touch on how and why smart contracts are necessary and some use cases of smart contracts. We will then describe Ethereum and its virtual machine, which have largely popularized smart contracts.
- Smart contract vulnerabilities (40 minutes): We will explore vulnerabilities that are typically exploited with smart contracts, and the very relevant consequences of some of these exploits. We will explore common vulnerabilities such as Reentrancy, which was responsible for TheDAO hack, Time Dependence, Transaction-Ordering Dependence and Critical Instruction Vulnerabilities.
- Vulnerability detection with deep learning (80 minutes): We will take a deep dive into how we can utilize learning techniques to automate the detection of smart contract vulnerabilities. We will use feature generation techniques to represent smart contracts and test under different deep learning models to compare the detection performance.
- Vulnerability Prevention (40 minutes): The tutorial will present an architecture based on Ethereum bytecode rewriting and validation for securing smart contracts in decentralized cryptocurrency systems.
- Applications (40 minutes): This section will discuss some useful blockchain applications, including applications with federated learning and IoT Data Management.

Total Duration: 4 hours
Some papers that will be referenced during the tutorial:

- Nakamoto, S. (2008). “Bitcoin: A Peer-to-Peer Electronic Cash System”. Bitcoin.org.
- Decker, S. & Wattenhofer, R. (2013). “Information Propagation in the Bitcoin Network” 13th IEEE International Conference on Peer-to-Peer Computing”
- Wang S. et al. (2019) “Blockchain-Enabled Smart Contracts: Architecture, Applications, and Future Trends”. IEEE Transactions on Systems, MAN, and Cybernetics.
- Sayeed S. et al. (2020) “Smart Contract: Attacks and Protections”. IEEE Access.
- Liao, J.-W., Tsai, T.-T., He, C.-K., & Tien, C.-W. (2019). “SoliAudit: Smart Contract Vulnerability Assessment Based on Machine Learning and Fuzz Testing”. Sixth International Conference on Internet of Things: Systems, Management and Security
- W. Wang, J. Song, G. Xu, Y. Li, H. Wang and C. Su (2020). "ContractWard: Automated Vulnerability Detection Models for Ethereum Smart Contracts," in IEEE Transactions on Network Science and Engineering.
- A. Tann, X. J. Han, S. S. Gupta, and Y.-S. Ong (2018). “Towards safer smart contracts: A sequence learning approach to detecting vulnerabilities”. arXiv preprint arXiv:1811.06632

- Gbadebo Ayoade, Erick Bauman, Latifur Khan, Kevin W. Hamlen.
  “Smart Contract Defense through Bytecode Rewriting”. Blockchain 2019, page 384-389
- Gbadebo Ayoade, Vishal Karande, Latifur Khan, Kevin W. Hamlen. “Decentralized IoT Data Management Using BlockChain and Trusted Execution Environment.” IEEE IRI 2018, page 15-22
- Sadaf MD Halim, Latifur Khan, Bhavani Thuraisingham (2020), “Next-Location Prediction Using Federated Learning on a Blockchain”. IEEE CogMI 2020.

Suitability for Virtual Presentation

This tutorial can work virtually as well as in person. If presented virtually, people will be able to easily follow along with the video stream.

Previous Tutorial Experiences (Related)

Dr. Latifur Khan has delivered the following invited talk on a related subject matter:

- “Automating Vulnerability Detection and Prevention in Smart Contracts” at The Second International Workshop on Blockchain and Data Management (BlockDM 2020), In Conjunction with ICDE 2020, April 20, 2020, Dallas, Texas, USA

Previous Tutorial Experiences (General)

Dr. Khan has conducted various tutorials over the course of his career. Some of these are listed below:

- Tutorial, “Secure IoT Stream Data Management and Analytics with Intel SGX,” *23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* KDD 2017, Halifax, Nova Scotia – Canada, August 13 - 17, 2017.
- Tutorial, *“IOT Big Stream Analytics,”* 26th International Joint Conference on Artificial Intelligence (IJCAI), Melbourne, Australia in August 2017.
- Tutorial, *“IOT Big stream Mining,” AAAI 2017 (The Thirty-First AAAI Conference on Artificial Intelligence will be held on February 4–9 at the Hilton San Francisco, San Francisco, California, USA.*
- Tutorial, *“IOT stream Mining,” SDM'17: the Seventeenth SIAM International Conference on Data Mining, Houston, Texas, USA, April 2017.*
- Tutorial, *“IOT stream Mining,” 22h ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* August 2016, San Francisco, USA.
- Tutorial, “Data Stream Mining and Its Applications”, *The 17th International Conference on Database Systems for Advanced Applications (DASFAA)*, April, 2012, Busan, South Korea.
- Half Day Tutorial, “Data Stream Mining Challenges and Techniques”, *The 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD),* Shenzhen, China, May 2011.

- Half Day Tutorial, “Matching Words and Pictures: Problems, and Applications in the Web*”, 2008 IEEE/WIC/ACM International Conference on Web Intelligence (WI-08),* December 2008, Sydney, Australia.
- Half Day Tutorial, “Matching Words and Pictures - Problems, Applications, and Progress,” *14th ACM International World Wide Web Conference, WWW2005*, May 2005, Chiba, Japan.
- Half Day Tutorial, “Matching Words and Pictures: Problems, Applications and Progress,” *ACM Fourteenth Conference on Information and Knowledge Management (CIKM),* November 2005, Bremen, Germany.

Related Tutorials: We could not find a tutorial in the last 2 years of ICC that directly relates to this content. In 2020, there was a tutorial titled “Blockchain technology and smart contracts in 5G networks and beyond" - this deals with how the industry can update its business process management using blockchain technology, whereas our tutorial deals with security concerns in blockchain technology. Tutorials in other venues have previously dealt with blockchains in general such as “On the privacy of transactions in account-based cryptocurrencies” at IEEE ICBC 2020. Some tutorials have also dealt specifically with smart contract vulnerabilities and their security previously, such as the talk titled “An Overview of Blockchain-Based Smart Contract Security Vulnerabilities”, at RSA 2018. While that tutorial gives an overview of the vulnerabilities it does not explain how to detect or prevent it. Our tutorial will explore the cutting edge of vulnerability detection and prediction.