Tong Ye
1998.08.20

Email : tongye@zju.edu.cn
Mobile : +86-18868102709

Education

- **ZheJiang University** — Hangzhou, China
  PhD candidate of Cyber Security — *Sep. 2020 – Now*
- **ZheJiang University** — Hangzhou, China
  Bachelor of Theoretical Physics; GPA: 3.86 — *Sep. 2016 – Jun. 2020*

Honorary award

- **Zhejiang Provincial Government Scholarship**
- **Outstanding Graduates of Zhejiang Province**

Papers

- **Tram: A Token-level Retrieval-augmented Mechanism for Source Code Summarization** 2022
  ACL under review
  Automatically generating human-readable text describing the functionality of a program is the intent of source code summarization. Although Neural Language Models (CodeBERT, GPT family) achieve significant performance in this field (Code $\longleftrightarrow$ Natural Language), an emerging trend is combining neural models with external knowledge. In this work, we explore a fine-grained token-level retrieval-augmented mechanism on the decoder side to help the vanilla neural model generate a better code summary.
- **State-of-the-Art Survey of Open-source Software Supply Chain Security** 2021
  Journal of Software *CCF-A*
  Software development is changing. Since the Internet allows far-flung development teams to collaboratively create software, open-source software supply chains are becoming more complex and sophisticated. This work tries to define the new open-source software supply chain model and presents a detailed survey of the security issues in the new open-source software supply chain architecture. Various emerging technologies, such as blockchain, machine learning (ML), and continuous fuzzing as solutions to the vulnerabilities in the open-source software supply chain have also been discussed.

Projects

- **Code Clone Detection**: Code clone refers to more than two duplicate or similar code fragments existing in a software system. We extract the Structural information (e.g., AST, CFG) of code to enrich the representation ability of code encoder, to align in the semantic space, and achieve the effect of the alignment of similar code in semantic space.
- **IOT Fuzzing**: Cyber attacks against IoT devices are a severe threat. These attacks exploit software vulnerabilities in IoT firmware. Fuzzing is an effective software testing technique for finding these vulnerabilities so they can be patched. We use the Function Code positioning of the protocol (privately) to improve the efficiency and coverage of fuzz.

Interests

- **NLP**: Text Generation, Code Summarization
- Math & Physics

Programming Skills

- **Languages**: C++, Python, Go