THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Clearblue: Towards building X-Ray for enterprise software

X-Ray: Knowing is believing

- X-Ray is a revolutionary technology underpin safety in the real world:
- Versatile in application scenarios:

Medical treatment

Science Research

Security Check

Archaeology

- Main reasons:
  - Input: the objects themselves without any intrusive preparation
  - Output: data representing the internal structure with no assumptions of downstream applications
- What's in my software?
  - Software more complex than other real-world objects
  - the duality of time and space: spatial code structure determines its computing process.
- What's shocking is that X-ray system for software does not exist
  - the input is the software's final state
  - the output is the materialized data of software behavior.

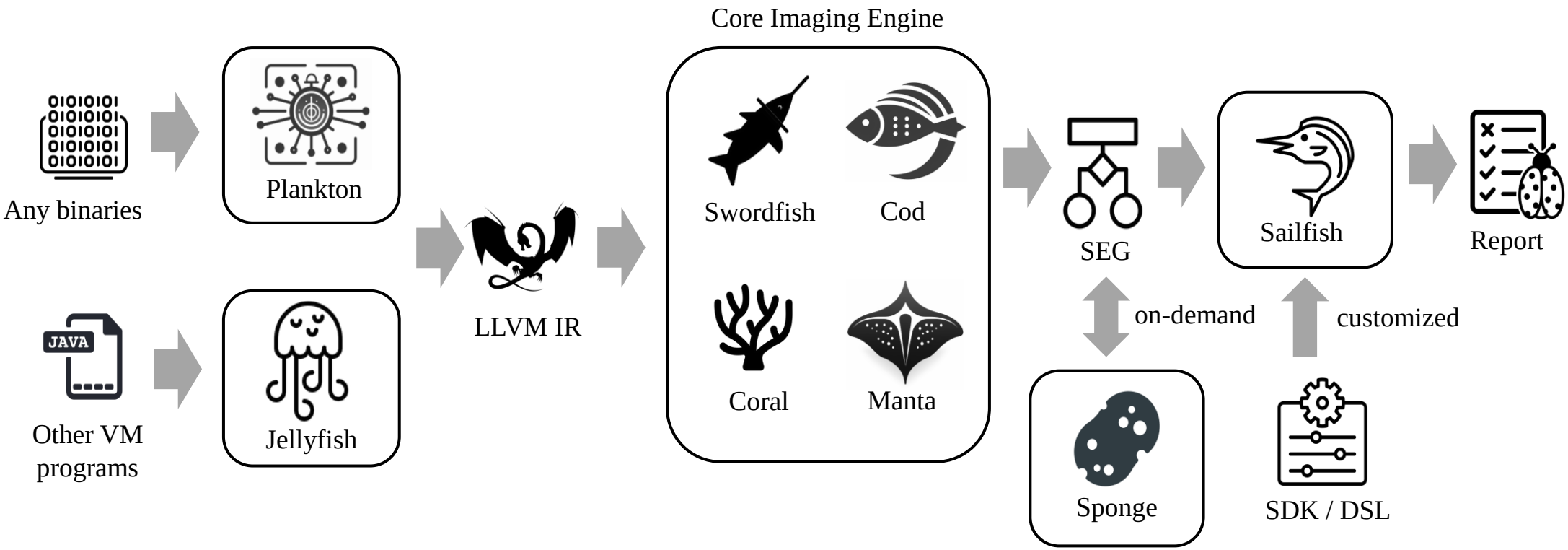
Modern Software

- There are over 100 million lines of code in your car.
- Large in scale and also “live” (constantly updated)
  - Linux kernel: 85,818 commits per year, 241 per day, 10 per hour, and one every 6 minutes.
- Enterprise software often involves a complex supply chain.
- Cloud-native (serverless + microservice) = programs small and simple + systems large and complex.
- The data for software behaviors are more than just for finding bugs.
  - Coding standards (MISRA, DO-178B), performance, security review, version management, dynamic testing, copyright verification, and many more.

Clearblue: Towards Building X-Ray System for Enterprise Software

- Provide the highest image resolution
  - Linux kernel (version 6.2), with nearly 30 million lines of code, 6 layers of inlining, and path-sensitivity
- Asynchronous analysis: separate analysis tasks from software data
  - Store software behavior data on disk (Linux kernel with 7G storage space)
  - Indexing of software behavior data to support on-demand queries
- No intrusion to the process of software construction and analyze binaries “out of box”
  - Supporting X86, ARM, Mach-O, Java bytecode
- A rich set of APIs based on Clearblue IR
  - Support both query of facts and searching of properties
  - Undergraduate students with basic compiler background: average time of writing CWE checkers is 30 minutes
- Begin to look like a basic database system

The System Architecture of Clearblue



A group of lovely marine creatures

Imaging Capabilities of Clearblue

- Main Conclusions
  - the only platform that fully supports asynchronous and precise analysis of Linux Kernel.
  - Clearblue offers the highest precision (inter-procedural, 6-level context-sensitive, fully path-sensitive).
    - SVF (3-layer context-sensitive, partially path-sensitive)
    - CSA (inter-procedural analysis within the file, file-level path-sensitive, context-sensitive)
    - Infer (context-sensitive, partially path sensitive, and explores up to 20 states)
    - CodeQL (not path-sensitive and context-sensitive)
  - Clearblue has much lower time and memory costs
  - In the past six months (July, 2024), Clearblue has found 74 CVEs and 400 security defects in much well-checked and well-established software, including close to 200 defects in the Linux Kernel.

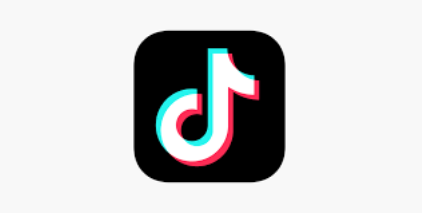
Clearblue in Huawei

HUAWEI

- First to win Huawei distinguished collaboration award two years in a roll
- Replace IDA Pro in Huawei software quality team
- Deployed in multiple product lines such as wireless and storage
- Clearblue function pointer analysis deployed in Huawei Hitest platform
- Clearblue concurrency analysis deployed in Huawei P4 platform
- Deployed in vulnerability detection of HarmonyOS

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Clearblue in Tiktok (on-going) :



- Supply chain analysis for Tiktok IOS App
  - Perform analysis of TikTok sensitive API, discovered two information leakage
  - Produce routine reports for TikTok sensitive API usage and sandbox change-impact-analysis
- Results well recognized by the Privacy & Security QA teams.
- Integrated into the program analysis technology system.

Clearblue in Ant Group

- Ant has difficulties using other tools for its sophisticated production environment.
- Clearblue’s direct binary analysis addresses problems such asNPD, UAF, Memory leak
- Clearblue discovered multiple crashing issues in IOT modules and OpenHarmony modules

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Thank you !

www.clearblueinnovations.org