CICI: UCSS: Secure Containers in High-Performance Computing Infrastructure

Overview

The security and privacy of high-performance computing (HPC) infrastructures are critically important as HPC infrastructures often process sensitive data, perform important scientific computations, and are relied upon by many organizations and individuals. Containers, as lightweight and isolated environments for running applications, e.g., Docker, and Singularity, are becoming increasingly popular in HPC infrastructures, while their security problems are paid less attention. Particularly, the containers in HPC infrastructure mainly face two security problems. **(i)** The container images are insecure. E.g., a recent study on 44 neuroscience container images of both Docker and Singularity shows that there are ***460 vulnerabilities per image***. **(ii)** The weak isolation may lead to vulnerabilities caused by running multiple containers on the same OS kernel. We observed ***11 such vulnerabilities since 2017***.

In this proposal, we aim to design secure containers for HPC infrastructures. **(i)** To address the insecure image problem, existing container image vulnerability scanners, e.g., Clair, Trivy, and Grype, face a *low coverage* challenge as they mainly conduct software version-based look-ups in public vulnerability databases. Therefore, in *Thrust 1*, we propose to design an efficient image vulnerability scanner with various innovative and feasible techniques, e.g., language-agnostic code representation with IR-reoptimization and natural language processing (Task 1-1), code similarity detection with graph neural network and triplet-loss network (Task 1-2), and scalable online search with locality-sensitive hashing (Task 1-3). **(ii)** Existing sandboxed runtime-based container solutions can address container isolation vulnerabilities by introducing a translation layer between the container and host kernel while facing low performance issues for HPC workloads. On the other hand, existing HPC container solutions (e.g., Docker, Singularity, Charliecloud, and Shifter) share the kernel with the host and are demonstrated to be vulnerable. Therefore, in *Thrust 2*, we propose to develop a secure and high-performance container runtime by using a lightweight virtual machine hypervisor (Task 2-1) with various customized optimizations for security and performance towards HPC workloads (Task 2-2), and dynamic image debloating to further remove attack surfaces (Task 2-3).

Intellectual Merit

Our project advances the security of containers in HPC infrastructure by offering a secure container platform for efficiently and securely running containers in HPC infrastructures. Our novel contributions are two-fold, i.e., an efficient image vulnerability scanner, and a secure and high-performance container runtime. *We envision this proposed work can significantly reduce the attack surfaces of containers in HPC infrastructures. Not limited, we aim to broaden the real impacts of this project by integrating our techniques and tools with existing HPC stakeholders to facilitate the security and privacy of HPC infrastructures.*

Broader Impacts

This research can have the following broader impacts. **(i)** The research will significantly advance the security of containers and thus the security of HPC infrastructures. The PIs plan to conduct system integration with HPC stakeholders; **(ii)** The PIs will integrate the proposed research into their curriculum development at both graduate and undergraduate levels. This proposed project will foster new research and educational opportunities at both the University of North Texas (UNT), a **Minority Serving Institution (MSI)** and a **Hispanic Serving Institution (HSI)**, and the University of Delaware (UD), which is in Delaware, an **EPSCoR** state in great need of education and research activities; **(iii)** The PIs will conduct outreach and educational activities in the K-12 community and promote the participation of students from underrepresented groups in Cybersecurity and HPC. **(iv)** In addition to disseminating the research results through publications, the PIs will also create a webpage and a publicly accessible GitHub repository to share the research results.