Carnegie Mellon University
Software Engineering Institute

Streamlining Containers for Edge Computing

Minimizing Container Size to Minimize Waste and Risk

SOFTWARE CONTAINERS HAVE BECOME A POPULAR STANDARD for packaging and distributing software at the tactical edge. But devices in edge environments are often limited, while containers are often larger and more resource-hungry than they need to be. To address the gap between what containers require and what edge environments offer, the SEI has developed the Container Minimization Tool (CMT) as part of its work supporting efficient and secure system deployment to the tactical edge. CMT automates the process of removing unused files and combining duplicate files to make containers work better in resource-limited environments.

Making Containers Less Demanding

Using Containers in Limited Environments

The DoD wants to use containers to support its vision of a cloud-to-edge continuum in which capabilities packaged as containers are pushed from the cloud to edge devices to support localized data processing. However, the tactical edge environment presents many challenges, including

- limits in storage space and computing power
- denied, degraded, intermittent, and limited-bandwidth (DDIL) networks
- high likelihood of bad actors trying to tamper with devices

Because container images need to be self-contained with all application dependencies, their use at the edge can clash with the constraints of edge environments. They are often significantly larger than they need to be, with much of their size wasted by unused or duplicated files. Such containers require greater transfer bandwidth and take a greater toll on device storage and the edge network. When the containers demand more size, weight, and power (SWaP) than the conditions of the edge can provide, new capabilities cannot be deployed. In addition, the larger the container is, the larger the number of vulnerabilities and consequently more attack surface is available for adversaries to exploit.

To address these challenges, the SEI created CMT, automated container minimization tool to minimize the storage size of a set of container images. This tool reduces storage waste without negatively impacting functionality and advances the state of the art in deduplication across container images.

A Greedy Algorithm That Prunes and Deduplicates

There are two main sources of storage waste in container images: unused files (such as development files) and duplicated files (identical files that are stored in different layers). The CMT addresses both types of waste.

At a high level, the CMT breaks up a set of container images into their individual files, reorganizes the layers, and reproduces a set of images and layers. This process prunes unnecessary files and deduplicates shared files from multiple images by creating a common container layer to hold them. While the size of the container is reduced, there is no impact to its functionality. The applications within the container run the same after minimization as they do before.

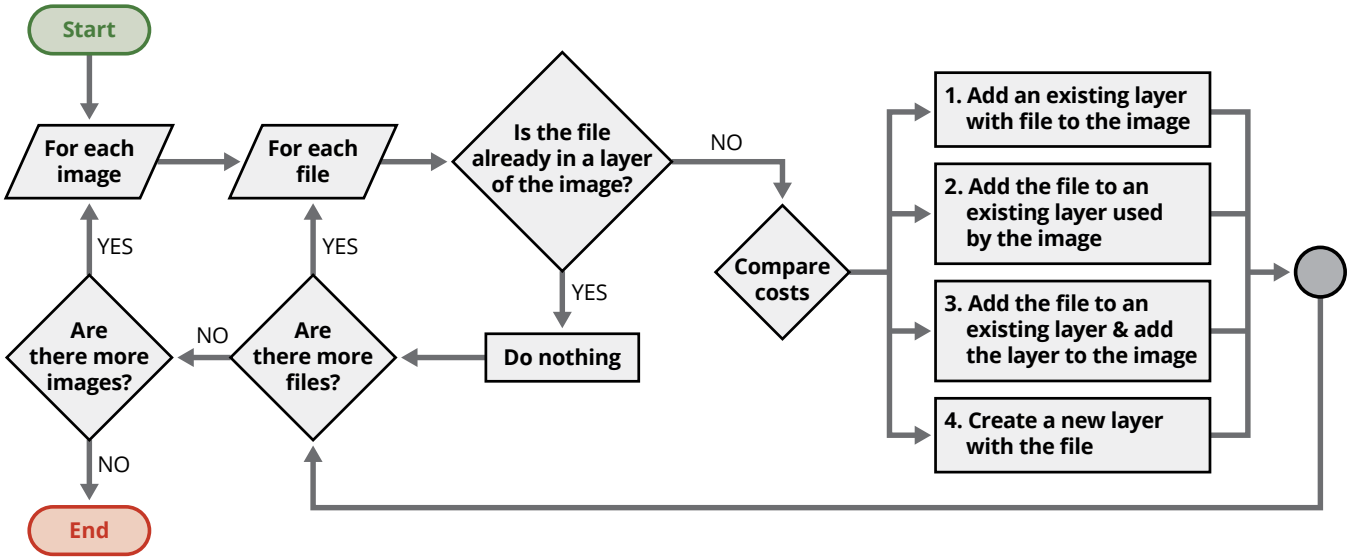

Figure 1: Flowchart depicting the process used by the CMT to streamline and deduplicate files in the container.

The CMT focuses on reducing three sources of cost: operational costs (the cost of too many layers), storage costs (the cost of duplicate files in layers), and network costs (the cost of large layers). The end result is that CMT reduces the storage and network costs of transferring container images from the cloud to the edge.

SEI testing indicates that the deduplication and pruning algorithms can reduce container image storage up to 10–30%. The CMT can run these algorithms quickly, processing 10 images with 225,000 files in approximately 51 minutes.

Learn More

Check out our presentation on automating container minimization for the edge: **resources.sei.cmu.edu/library/asset-view.cfm?assetid=889345**

Learn more about the SEI's work in edge computing: **sei.cmu.edu/our-work/edge-computing/index.cfm**

Get Started Today

The CMT has the potential to allow DoD organizations to field more capability per SWaP at faster deployment speeds while reducing the number of software vulnerabilities that may be present in unused files.

Contact us today to learn how the CMT can help your team deploy more efficiently and reliably in edge environments.

About the SEI

Always focused on the future, the Software Engineering Institute (SEI) advances software as a strategic advantage for national security. We lead research and direct transition of software engineering, cybersecurity, and artificial intelligence technologies at the intersection of academia, industry, and government. We serve the nation as a federally funded research and development center (FFRDC) sponsored by the U.S. Department of Defense (DoD) and are based at Carnegie Mellon University, a global research university annually rated among the best for its programs in computer science and engineering.

Contact Us

CARNEGIE MELLON UNIVERSITY
SOFTWARE ENGINEERING INSTITUTE
4500 FIFTH AVENUE; PITTSBURGH, PA 15213-2612

sei.cmu.edu
412.268.5800 | 888.201.4479
info@sei.cmu.edu