SpaceOS: Secure and Efficient Cloud-to-Edge Computing with Type-Safe Unikernels

Thomas Gazagnaire thomas@tarides.com
KC Sivaramakrishnan kc@tarides.com
Miklos Tomka miklos@tarides.com

Introduction

Over the last decades, one of the European Union's challenges and policies was to develop a strong and vibrant economic base by embracing digital transformation. On the business side, this was translated into a global digitalisation process of business infrastructures, primarily driven by the three historical cloud platforms based in the US: Amazon EC2, Google Cloud, and Microsoft Azure. It allowed companies and users to temporarily rent digital resources and use them with a lot of flexibility. This was made possible thanks to disruptive virtualisation technologies (that Tarides' founders helped build while working at XenSource and Citrix) to secure multi-tenant usage of underutilised data centres. It profoundly changed the way the industry built and used its software. It also created new markets as entire new segments of industries went digital. As a result, the European digital transformation is now mainly a synonym for moving data and processing power to the Cloud, mostly controlled by US entities.

In the last decade, new systems have been extending the existing cloud-based infrastructure to create what we now call the "Edge". This was possible thanks to tremendous advances in mobile and sensing technologies and to the ongoing rapid deployment of "smart" digital infrastructure that augments the physical environment to form the so-called "Internet of Things" (IoT). However, this infrastructure usually relies on a centralised cloud, transmitting vast amounts of data to and from physically remote environments (sometimes in a different country or continent). Private companies are driving these systems' rapid development and deployment and are pushing to reduce time-to-market by iterating quickly to find a market fit. Because of this, security and resource efficiency has usually taken a back seat to convenience. The cost for this is data insecurity, resource inefficiencies, high response latency and unpredictable reliability of services.

With the rise of Machine Learning on the Edge and "smart" IoT devices, the demand for highly-interconnected devices is only increasing. Unfortunately, the security of these devices remains unchecked and makes them susceptible to security vulnerabilities, leaving consumers and businesses open to exploitation. For instance, there is evidence that Amazon's Alexa device is sending audio recordings without users' consent and in 2019, the "smart" city of Baltimore was held hostage by hackers' ransomware. Moreover, deploying millions of sensors and computing nodes will further increase energy consumption, with research suggesting that 3.5% of global emissions will be produced by IoT devices by 2025. Unfortunately, the IoT industry has yet to provide resource-efficiency solutions to reduce this contribution to climate change.

Fortunately, the EU is now advocating more actively for an improved focus on the security of these systems, for instance with the Cyber-Resilience Act. To help companies solve these issues, Tarides has created SpaceOS: a disruptive platform for managing digital IoT/Edge infrastructure at scale, securely and efficiently. It combines hardware and software elements that invert the current cloud-centric model by building an operating system designed to securely connect physical spaces with extremely low latency, high bandwidth local-area computation capabilities and service discovery. It allows operators to turn a fleet of captors, sensors, computing nodes and network elements (for instance running in a remote farm or a satellite constellation) into an autonomous data centre, where computing resources can be tracked efficiently and temporarily rented to users to provide "Platform-as-a Service". This turns any vertical IoT deployment (a constellation of satellites, a smart city, etc.) into an autonomous, private cloud, allowing better utilisation of local resources and improved security. This grants, in turn, the local deployment of a SaaS ("Software-as-a-Service") platform, where developers can create new dedicated applications. At the same time, users can easily install these applications on local digital resources.

Motivation

Two decades ago, a major technological breakthrough, so-called "virtualisation", allowed multiple (unmodified) operating systems to run simultaneously on the same hardware without overhead. Virtualisation allowed each virtual machine to be securely isolated, allowing companies with large, under-utilised data centres to consolidate their workloads and rent their internal infrastructure to external customers. By doing so, these companies (such as Amazon, Google and Microsoft) have thrived by creating an entirely new market, which has become the centre of the digital revolution: the public cloud.

In the interim, the design and use of the cloud have become more specialised and focused. The logical unit of computation is now more aligned with customers' core-business logic. It covers a full range of uses, including complete virtual machines (via Infrastructure-as-a-Service), application environments (via Platform-as-a-Service), and even raw functions (with Serverless and Function-as-a-Service). This abstracts users from requiring detailed knowledge of the full software stack, and they can focus on their business logic. However, despite this specialisation and customer-focused development, the deployment story for these logical units is still the same and has not been specialised in an equivalent or complementary manner. So, while the logical units from the Cloud are becoming leaner, the supporting software stack is becoming increasingly complex. This complexity is causing a larger attack surface, increasing the risk of cyberattacks, and is a waste of computing resources.

Tarides sees this increased complexity as an opportunity. The SpaceOS platform is built upon major innovations in programming languages, operating system security and distributed systems, initiated by Tarides' founders more than 10 years ago at the University of Cambridge (UK) and Inria. These technology breakthroughs, specifically the concept of "unikernels", make it possible to radically simplify how applications are built and deployed for the cloud. Unikernels allow applications to be specialised to their development environment, generating applications of size roughly 4% of that of similar systems[1], at least 70% more secure[2] and 10 times more resource-efficient for cloud and IoT deployments[3]. The latter suffers from security and resource efficiency issues arguably more than cloud deployments. Fortunately, as more connected device vendors begin to support virtualisation technologies, IoT deployments will directly benefit from adopting technologies like unikernels. Software fault isolation techniques (such as WebAssembly compilation) can be used for devices not supporting virtualisation. We also envision having dedicated hardware to ensure this isolation based on RISC-V open-hardware specifications.

Combining unikernel technologies and hardware/software fault isolation techniques is the missing innovation that will consolidate the fragmented Cloud-to-Edge market.

[1] https://www.usenix.org/system/files/conference/usenixsecurity15/sec15-paper-kaloper-mersinjak.pdf

[2] https://msrc-blog.microsoft.com/2019/07/18/we-need-a-safer-systems-programming-language/

[3] https://www.lemonde.fr/pixels/article/2019/12/30/a-leipzig-hackers-et-militants-pour-le-climat-font-front-commun_6024362_4408996.html

Current Status

Tarides has developed and used unikernel applications for Cloud applications for many years. We are now focusing on using unikernels on embedded devices.
The two main technical components that we are currently developing are (1) an SDK to write secure-by-design applications built around the OCaml programming language and (2) the MirageOS unikernel project to specialise applications to their runtime environment. We aim to package these technologies into SpaceOS, a secure, flexible and easy-to-use platform for the Cloud-to-Edge continuum, initially focused on the Satellite and Ground Station industries.

OCaml

OCaml is an open-source, type-safe, functional programming language allowing developers to write safer applications. Based on research developed for decades in programming languages at Inria and the University of Cambridge, the OCaml compiler analyses programs to automatically detect the most common security issues (memory and type safety) before the program even runs. This enables a “security-by-design” approach to software development. The OCaml compiler has been developed for over 20 years at Inria (Institut National de Recherche en Informatique et Automatique). Xavier Leroy[4] created the OCaml language, received numerous awards and obtained the first chair of software engineering sciences at “College de France”.

Today, OCaml is the language of choice for building major verification software tools (Coq, Why3, AltErgo, Imandra), with use extending from pure research into industry. Tarides is heavily involved in developing and maintaining the OCaml language and aims to incorporate formal methods within the industry's increasingly complex systems. Tarides’ work also directly impacts learning and teaching within computer and data science. Following the release of Real World OCaml in 2013 (co-authored by Tarides’ co-founder), the book has been adopted in many other universities as a reference textbook for various computer science courses, including at Cornell, Harvard and Princeton.

MirageOS

Tarides co-founders developed the concept of unikernels in 2015 at the University of Cambridge. Unikernels are small, potentially transient computer modules specialised to undertake a single task at the point in time when it is needed. Because of their reduced size, they are far more secure than traditional operating systems and can be started up and shut down quickly and cheaply, providing flexibility and additional security.

MirageOS[5] is an open-source project which implements unikernels written in OCaml. It splits a traditional operating system into a collection of flexible, reusable, type-safe system libraries that can be reused in various contexts to build secure-by-design applications that run directly on a hypervisor (such as Xen) or bare metal machines. MirageOS provides many benefits compared to a traditional OS; MirageOS unikernels are small, simple and quick, using just enough code to enable the relevant application or process to run (about 4% of a traditional operating system[6]). As a result, MirageOS applications have improved security and smaller resource footprints[7].

This research and technology breakthrough received multiple awards (including 2016’s Cambridge Computer Lab Ring’s “Best Company” and “Best Paper” awards[8]). In addition, it led to the creation in

[4] https://en.wikipedia.org/wiki/Xavier_Leroy

[5] https://mirage.io

[6] https://nqsb.io/nqsbtls-usenix-security15.pdf

[7] https://mirage.io/blog/ccc-2019-leipzig , Originally published by Le Monde on December 30th, 2019: https://www.lemonde.fr/pixels/article/2019/12/30/a-leipzig-hackers-et-militants-pour-le-climat-font-front-commun_6024362_4408996.html

[8] https://www.cst.cam.ac.uk/ring/awards

2015 of the “Unikernel Systems” company (acquired by Docker in 2016) and then, in 2018, of the Tarides company.

SpaceOS

SpaceOS aims to be a groundbreaking operating system tailored for the space industry vertical, designed to meet the unique demands of the NewSpace era - which is a good example of the Cloud-Edge Continuum in action: historically, data is generated on satellites, with poor connectivity, low bandwidth and high-latency and is processed on ground station with powerful computation powers. It takes, on average, 12h for images to be sent to ground stations, and bandwidth is now the limiting factor with the availability of new-generation sensors (like Hyperspectral cameras). As a result, the NewSpace is trying to move computation closer to the data, thus opening satellites to new attack surfaces. Hence, key features of SpaceOS include multi-tenancy, which ensures strict separation and isolation of software payloads; a flexible software platform that offers a fast, efficient, and robust solution with a small footprint for quick transfers and seamless updates; scalable security that employs a memory-safe language, cryptographic libraries, and formal verification for the maximum protection; and ease-of-use, characterised by a user-friendly SDK, comprehensive documentation, and compatibility with industry-standard development tools for effortless implementation.

In 2023, Tarides demonstrated a proof of concept of SpaceOS to Thales Alenia Space. SpaceOS was selected to provide onboard AI image analysis software for Thales satellites. The goal was to achieve identical functional results and flexibility as the Linux/Docker benchmark while demonstrating improvements in efficiency and cybersecurity. The results showed that the PoC successfully achieved a 100% replication of the benchmark's functionality while delivering significant benefits, including a 20x reduction in executable code size, a 2.5x decrease in memory requirements, a 20% improvement in performance, enhanced security, and simplified software deployment and updates. While still very early, SpaceOS has the potential as a pioneering operating system for the space industry, offering a secure, efficient, and adaptable solution that addresses the challenges and requirements of NewSpace.

Research Challenges

MirageOS unikernels enable applications to be specialised to their runtime environments, whether they are deployed on cloud-based hypervisors or resource-constrained edge devices. While deploying these technologies on the cloud has been achieved by various companies and entities in the past, adapting and improving cloud-based technology for embedded devices presents several research challenges.

- **Cybersecurity**: Ensuring high levels of cybersecurity in embedded systems is crucial. Developing a secure-by-design approach, measurable security benchmarks, and techniques to maintain security in increasingly complex environments will be essential.
- **Resource Utilization**: Efficiently utilising system resources is a critical challenge, particularly in resource-constrained edge devices. Minimising the footprint of existing applications and reducing the need for hardware upgrades when deploying new, more complex applications is a key research area.
- **Virtualisation and Isolation**: Relying on virtualisation is beneficial when available, but it can be challenging for embedded processors where specific hardware extensions might not be present. Various software fault isolation techniques, such as sandboxing, Native Client (NaCl), and WebAssembly (Wasm), can provide isolation and protection.
- **Secure and Flexible Upgrades**: Creating flexible and powerful upgrade capabilities for seamless remote device updates while ensuring the security of these updates is a significant

challenge. It is crucial to develop methods to verify the authenticity and integrity of updates and safely deploy them without disrupting the system.

- **Hardware and Platform Compatibility**: Another research challenge is designing "hardware as a service" solutions that cater to diverse needs. Developing simulators to test new applications on an extensive range of platforms and enabling "digital twin"-style testing environments will help ensure compatibility and performance across different hardware and deployment scenarios.

Addressing these research challenges will help unlock the full potential of SpaceOS as a combination of OCaml and MirageOS, paving the way for secure, efficient, and adaptable solutions that can be effectively deployed in embedded devices within the space industry and beyond.

Why Tarides?

Tarides is a French company committed to advancing operating systems and related IoT and edge computing software development. Our team of researchers, including a third holding PhDs and leading the charge in computer science research, collaborates to create cutting-edge open-source technology based on the OCaml language and the MirageOS unikernel project. OCaml was invented and is still maintained by a team primarily based in France. MirageOS was created in Cambridge and maintained mainly by UK, France and German teams. Thus our solutions aim to provide a world-class alternative for European companies seeking to avoid reliance on technologies controlled by global technology giants like Apple, Microsoft, and Google.

By fostering a collaborative research environment and leveraging an open-source model, Tarides invites contributions from experts worldwide while retaining a strong European identity. With SpaceOS, we aim to create and maintain a high-quality European IoT and edge computing alternative with OCaml and MirageOS at its core. This aligns perfectly with the EU's call for innovation in the cloud-to-edge continuum, emphasising the importance of open source, open standards, and technological sovereignty.

This world-class technology, developed entirely in Europe, seeks support for the required research to advance its objectives. Thomas Gazagnaire of Tarides, having previously developed cloud technology acquired by Docker, is now dedicated to retaining this innovation within Europe. In anticipation of potential challenges and risks, Tarides will employ strategies to mitigate or overcome obstacles, ensuring the region's independence in a world increasingly reliant on IoT, smart devices, and New Space technologies, where every individual and company is connected.