Time-accurate Network Simulation Interconnecting VMs toward stealth analysis

Executive summary: The proposed work aims at designing an evaluation environment for distributed infrastructures where the instances of the real application are executed in full QEMU-based virtual machines, interconnected by the SimGrid network simulator. Starting from a working prototype, the proposed work will enable the study of malicious applications.

Advisors:

- Martin Quinson (ENS-Rennes, IRISA, team Myriads) `Martin.Quinson@ens-rennes.fr`
- Louis Rilling (DGA, team Myriads) `Louis.Rilling@irisa.fr`
- Matthieu Simonin (Inria Rennes) `Matthieu.Simonin@inria.fr`

Team: Myriads. **Laboratory:** IRISA, Rennes (head: Guillaume Gravier – guig@irisa.fr).

Key skills: Deep understanding of OSes, Networks and VMs; System Programming on Linux.

Context and Description

By nature, distributed applications are challenging to analyze and debug because of their size, complexity and dynamicity. These characteristics make it very difficult to actually test the systems in a reliable and reproducible manner. Simulation constitutes an appealing alternative toward convenient experiments, but often either requires to reimplement the target application using a specific interface, or mandates intrusive inspection techniques.

QEMU is a whole system emulator that can be used as a virtual machine to run the differing nodes of a distributed infrastructure on a single machine. The Panda project builds upon QEMU for the reverse engineering and live analysis of arbitrary systems. SimGrid is a simulator of distributed applications in heterogeneous distributed environments. Its key features are its sound performance models (enabling accurate performance prediction in non-trivial scenarios) as well as its ability to run directly legacy applications written with the MPI standard.

This internship aims at leveraging the predictive power of SimGrid on unmodified, non-trivial distributed applications executed within modified QEMU-based virtual machines. The ultimate goal is to enable to study codes that detect and evade any analysis. To this end, we want to extend the stealthiness of existing analysis frameworks by completely simulating the network.

Detailed Work Plan

The intern will be provided with a working prototype that can run simple applications between QEMU VMs through SimGrid, opening many research directions. The intern is expected to choose and complete two to three work leads from the four ones provided here:

- *Framework validation on real applications.* The intern will validate the prototype on several real distributed applications of increasing complexity, to gradually stress the implemented mechanisms. The proposed applications are ShareLaTeX, Ceph and DPDK.
- *Performance evaluation to understand the tool practical usability.* QEMU is currently used in full emulation mode, that comes with a high performance penalty. The intern will quantify this performance loss for several applications used out of the box.
- *Overall approach portability beyond the current implementation.* We want to explore how our work can be generalized. We want to remove the need of full emulation (that hinders performance) to allow hardware-assisted virtualization. Providing sufficient control over the time in this case will require new techniques, as the existing ones heavily depend on the full emulation. Similarly, we want to explore how our work could be generalized and adapted to KVM and Xen.
- *Evaluating the framework intrusiveness on sandboxes' analysis tools.* Finally, we want to evaluate the interference induced by our module once integrated to open-source analysis frameworks such as Panda or Drakvuf. In particular, the strict control of time achieved by our module should not hinder the sandbox' analysis power.

Bibliography

- H. Casanova, A. Giersch, A. Legrand, M. Quinson and F. Suter. *Versatile, Scalable, and Accurate Simulation of Distributed Applications and Platforms*, Journal of Parallel and Distributed Computing 74(10), 2014. http://hal.inria.fr/hal-01017319.
- B. Dolan-Gavitt, J. Hodosh, P. Hulin, T. Leek, R. Whelan. *Repeatable Reverse Engineering with PANDA*. 5th Program Protection and Reverse Engineering Workshop, Los Angeles, California, December 2015. https://apps.dtic.mil/sti/pdfs/AD1034415.pdf
- T. Lengyel, S. Maresca, B. Payne, G. Webster, S. Vogl and A. Kiayias. *Scalability, Fidelity and Stealth in the DRAKVUF Dynamic Malware Analysis System*. Proceedings of the 30th Annual Computer Security Applications Conference, 2014.
- J. Mirkovic, T. Benzel, T. Faber, R. Braden, J. Wroclawski and S. Schwab. *The DETER Project: Advancing the Science of Cyber Security Experimentation and Test.* In Proceedings of the IEEE Technologies for Homeland Security Conference 2010 (HST'10).
- N. Miramirkhani, M. Priya Appini, N. Nikiforakis, M. Polychronakis. *Spotless Sandboxes: Evading Malware Analysis Systems using Wear-and-Tear Artifacts.* In Proceedings of the 2017 IEEE Symposium on Security and Privacy. https://securitee.org/files/wearntear-oakland2017.pdf
- H. Tazaki, F. Urbani, E. Mancini, M. Lacage, D. Camara, T. Turletti, and W. Dabbous, *Direct Code Execution: Revisiting Library OS Architecture for Reproducible Network Experiments*. In the 9th International Conference on emerging Networking EXperiments and Technologies (CoNEXT'13). https://hal.inria.fr/hal-00880870