Using software defined networks for computing with Protected Health Information (PHI)

Evan Bollig, Jeff McDonald, and J. Vinals. Minnesota Supercomputing Institute, University of Minnesota.

The application that we are considering involves providing supercomputing resources in a secure manner to manipulate or analyze protected health information that, while complying with HIPAA requirements, it does not introduce excessive overhead in the operation of a general purpose supercomputing infrastructure. As is typical in a University environment, the largest fraction of computation or storage is not subject to specific privacy guidelines, yet medical schools or university hospital could better profit from the infrastructure in place if it were HIPAA compliant.

The case that we would like to develop further involves analyzing large surgical note collections using natural language processing (NLP) techniques. This requires computational resources that greatly exceed those typically available in academic health centers, certainly at the University of Minnesota. External high performance computing centers such as the Minnesota Supercomputing Institute (MSI) have sufficient resources; however, strict HIPAA compliance requirements limit their use for processing Protected Health Information (PHI). We have developed a protocol for using our local HPC resources in a secure manner, and they are currently being used to analyze clinical notes from our hospitals. We would like to experiment with all aspects of the protocol in a cloud environment that is also capable of producing supercomputer class performance, so as to eventually deploy it remotely.

Our currently approved protocol calls for encryption of all data transfers between hospitals and MSI, confining all analyses to volatile memory, node cluster isolation and termination protocols under the detection of any other attempted access, and scrubbing each compute node at the end of the taks. The entire process is remotely driven by internal hospital servers which control provisioning, scheduling and control. Our current pipeline scales up to 512 client agents running on the HPC system, with the limitation being the throughput of the messaging server in the hospital.

We would like to experiment with the NSF cloud system to replace this protocol with a more general purpose one benefiting many nascent users who would like to explore the power of a sizable compute resource without incurring the overhead of providing the staff, power and cooling needed for a bare metal system. MSI would engage the NSF Cloud offerings to provide two key functions for the PHI workflows.

First, the SSH tunnels are currently used to provide a secure, compliant connection to storage at the user's site (in this case, the University of Minnesota Academic Health Center) could be replaced with a Software Defined Network (SDN) that would extend into the space of the user.

The SDN would effectively create a private cloud between NSFCloud resources and the user's site.

Second, the bare-metal components of our current compute system would be replaced with a virtual cluster appropriately sized to meet the needs of each user. Connecting the virtual cluster with the private cloud network would create a virtual private cloud of resources for this user community.

MSI is well qualified to provide this and similar real world use cases to develop, and also to provide some expertise on the compliance side. We feel this solution would be of great benefit to clinical sites needing computational resources but without having any of those resources locally available. In addition, the availability of proven PHI/HIPPA certified solutions would give University security officers confidence in allowing their campus to use the NSFCloud solution for PHI analytics.