EE/CprE/SE 491 SDMAY-19

Week 3 Report

9/1/2018 – 9/8/2018

Group number: ***5***

Project title: ***Cyber Network Capture Generator***

Client : ***Argonne National Laboratory***

Advisor : ***Benjamin Blakely***

Team Members:

Jacob Perin - ***Scribe***

Luke Tang - ***Meeting Facilitator***

Collin McElvain - ***Chief Architect***

Abdelrahman Baz - ***Chief Architect***

Hazem Abdeltawab - ***Test Manager***

Bernard Ang - ***Report Manager***

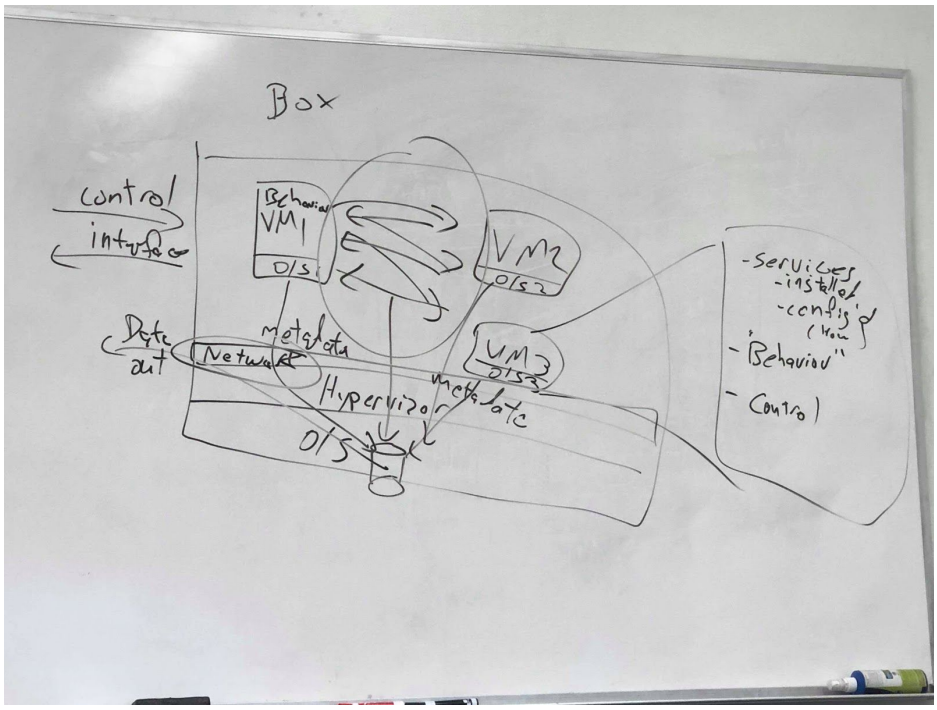
Weekly Summary

This week,we had our first meeting as a group to to discuss on the roles we will be having in the team. Also, we met with the client together with the faculty advisor to introduce ourselves as well as to learn and understand more about the project. After that, we as a team decided that we should research more on the tools that may be used to complete the project. Thus, our objective for the week was to learn more about the tools and talk about it on the next meeting so we are able to come into a conclusion of what tools we will be using for the projects. Each members are assigned with two different tools to research on. As of now, no changes has yet to be made to the project.

Past week accomplishments

As a team, we had a meeting with the client and the faculty advisor on Wednesday, 5th of September at 12PM in Room 3138, Coover Hall. In the meeting, we discussed about several different topics related to the project such as the expected requirements, the deliverables required throughout the project, visualizing the outcome and the flow of the project (Figure 1 below) as well as the different tools that we are recommended to use for the project. Each team members

were assigned to 2 different tools to research on and the outcome is as shown below :



- • Bernard :
  - Xen
    - To run the Virtual machines required for our project, we require a hypervisor. Xen is a hypervisor that uses microkernels to provide services that allow multiple operating systems to run on the same computer hardware. The main reason that Xen is a good choice as our hypervisor for the project is that the Xen project is fully free and open sourced and it is one of the requirements from our client. Also, Xen has good scalability compared to the other hypervisor choices. Xen also does well in separating the hypervisor execution from management OS,management stack,device drivers and it's guests(components) which makes it more efficient.

  - Puppet
    - To configure our virtual machines properly, we require a configuration management tool. Puppet is a free and fully open sourced configuration management tool that can be used to deploy,manage,configure and maintain a server machine. Especially for our project, puppet is very helpful as it can be used to manage machines that run on a large number of platforms such as MACOS, Redhat, Windows and many more. Also,

Puppet is a popular choice, being used by big companies such as Google and Redhat. Other than that, Puppet also works smoothly even when deployed in a large infrastructure, which is an advantage for us as the completed project may be used and further developed by the Argonne National Lab in the future.

- • Jacob :
  - Xen
    - Why do we need XEN?
      - A large portion of our project is creating an environment that will allow a user to create guest OS(s) on the fly and capture traffic between them. We will need virtualization software to accomplish the task of creating the necessary guests.
    - Xen is a bare-metal hypervisor that allows a user to run virtual machines
    - Supports many of the requirements described by our client, including:
      - Supports windows and many flavors of linux
      - Supports the cloud (probably not relevant to us, but client wanted option for the future)
      - Can run on a large range of hardware, due to feature "para-virtualization" (we do not know what client has to run the hypervisor -- on a budget)
    - Tools that allow for more extensive features:
      - XAPI
        - API toolstack/backend for managing application Guest OS remotely (e.g. upload malware, custom setup, etc.)
      - XEN VCN XVP
        - Manages VNC connections
      - XenCenter (or OpenXenManager)
        - Resource pool management
  - PCAP
    - What is PCAP?
      - Packet Capture Format
    - Why we need to understand this?
      - Storing full packet capture is extensive. Even a short period of logging data will be costly in terms of storage. We will need to do extensive filtering and testing of that filtering to ensure we only record what is absolutely necessary.
      - Our application will have to have to option for different types of network capture to remove such problematic scenarios. Some of our options include: full packet capture, augmented flow, and

network flow. Which the client wants, how we configure this into the application, and other details are most likely going to depend on the application.

- • Collin :
  - ➢ Chef
    - ■ Chef is an open source configuration management software that will help in the automation of our environment. Chef is made up of different cookbooks and recipes that allow for system configuration in either on-premise of virtual environments. Chef seems to be one of the most flexible configuration management tools out there. This will be incredibly useful in the setup of our project's environment as hopefully it will help automate the system setup process for each test run.
    - ■ Chef will utilizes a test-and-repair system for changes. This means that chef will only make changes to our system if one of the systems diverges from its assigned recipe/resource.
    - ■ Chef also includes many resources for learning. This includes ChefRally, a tutorial/module based learning program for free.
  - ➢ Expect Scripts
    - ■ Expect is a scripting language used to automate a conversation between multiple systems. Expect requires the user to write the request and response for all systems involved. This is a useful tool for long checks on systems. However, the amount of code required may not be useful in our project. The only case I see this being useful is maybe in checking system status' or other repeated system conversations.
    - ■ Expect is already an available package on most linux systems.
- • Lucas:
  - ➢ Virtualization: KVM (Kernel-based Virtual Machine)
    - ■ KVM is a virtualization infrastructure of type Hypervisor that operates like Unix for the linux kernel. KVM itself does not perform emulation, but rather exposes /dev/kvm interface that allows simulated I/O, map video display to host system, and set up VM address space where a custom BIOS can bootstrap into the main OS. It is a full virtualization solution for Linux on x86 hardware with virtualization extensions. KVM would be useful specific to this project because it can run multiple virtual machines running Linux or Windows images. Each machine would have its own private hardware such as network card, disk, graphics adapter, and more. It is also open source software.
  - ➢ Selenium
    - ■ Selenium is an open source cross-platform software testing framework for

web applications. Selenium is a large project with many components: Selenium IDE, Selenium client API, Selenium WebDriver, Selenium Remote Control, and Selenium Grid. The Integrated Development Environment (IDE) is used for testing as a Firefox Add-On, as well as with a recent integration with Chrome. The outdated client API can be used to write tests with calls to the Selenium Client API. The Selenium WebDriver accepts commands and sends them to a browser and retrieves the results, including the execution of JavaScript commands. Selenium Remote Control is a Java server that accepts HTTP commands for the browser to write automated tests for web applications in any programming language. Selenium Grid is a server that allows tests to be run for web browser instances running on remote machines. Overall Selenium is an excellent WebDriver framework that can be used to emulate real human traffic across several instances in virtualization for this project.

- •Abdelrahman:
  - Netflow
    - To capture the network traffic between the virtual machines, we need a tool like Netflow. Netflow is a protocol developed by Cisco to collect and record all IP traffic going to and from a Cisco router that is netflow enabled. After the traffic has been recorded, a netflow analyzer is used to organize the traffic and analyze it. The protocol identifies the the source and destination of each packet, the source and destination ports for TCP or UDP, class of service, layer 3 protocol type, and router or switch interface. Then it groups packets that have the same attributes into a flow. Netflow and other traffic capturing tools could cause too much overhead on the CPU when when expanding the flow tuple to include more details, which is an important point to be considered when choosing an efficient tool.

  - Puppet
    - Puppet is a configuration management technology to manage the infrastructure on physical or virtual machines. It is an open-source software configuration management tool developed using Ruby which helps in managing complex infrastructure on the fly. We need to use a tool like Puppet to automate the configuration of the virtual machines. The way Puppet works is that the user describes system resources and their state using Puppet's declarative language or a Ruby DSL. The description is stored in files called "Puppet manifests". Puppet, then, discovers the system information via a utility called Facter, and compiles the Puppet

manifests into a system-specific catalog containing resources and resource dependency, which are applied against the target systems. Any actions taken by Puppet are then reported.

- • Hazem :
  - VirtualBox
    - VirtualBox is a free virtualization software done by Sun Microsystems (now Oracle) used for the purpose of creating virtual machines running derivatives or versions of Microsoft, Linux, BSD, Solaris and others. It is regarded as one of the most used virtualization softwares in the world, along with VMware. However, VirtualBox has some very serious limitations. For example, it has a very low transfer rate to and from USB2 devices and most of the important functions are only available under a commercial license, thus making it harder to finish the required work.
  - Netflow
    - Netflow is a software developed by Cisco that helps in recording packets as they enter and leave a certain interface. Netflow is beneficial because it allows a System Administration to access starting destinations and final destinations of packers, which helps in creating a more securely advanced interface for the project currently in-progress. In addition, there are other softwares that work like Netflow; However, most of them generate more output than necessary, thereby making the results harder to parse. That is why to choose Netflow, and that it runs low on the CPU memory.

Individual contributions

Team member	Contribution	Weekly Hours	Cumulative Hours
Bernard Ang	Researched on Xen and Puppet	**7**	**7**
Collin Mcelvain	Researched Chef and Expect scripts	**5**	**5**
Jacob Perin	Researched Xen and PCAP	**6**	**6**
Lucas Tang	Researched Selenium and KVM	**6**	**6**
Abdelrahman Baz	Researched Netflow and Puppet	**6**	**6**

Hazem Abdeltawab	VirtualBox and Netflow	**15**	**15**

Plan to accomplish for the next week

- Bernard
  - Continuing research
    - Further follow up on my research on Puppet, understanding how it may work and how to deploy it.
    - Continue research on Xen and try to run a few machines on it.
- Hazem Abdeltawab
  - Continuing research
    - Further follow up on VirtualBox, adding any new information as necessary
    - Continue research on Netflow and other related softwares.
- Jacob Perin
  - Continue Research
    - Initial testing on Xen usability (sharp learning curve)
    - Install / use tools discussed, and determine if usable in our application
  - Collaborate with Team members
    - Discuss evaluated pros/cons of Xen and why I believe we should use this virtualization software versus other tools available
    - Create Diagram with tools + our idea of system for client meeting on Wednesday Morning.
- Collin McElvain
  - Continue Research
    - Begin experimenting with Chef.
  - Architecture
    - Begin assembling the researched software's and start creating a architecture for this project (use case diagrams, component diagrams, etc.).

- Abdelrahman Baz
  - Continue Research
    - Begin experimenting with Puppet and Netflow .
    - Read about how VirtualBox works and how it creates virtual machines.