THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

COLLEGE OF INFORMATION SCIENCES AND TECHNOLOGY

Polymorphic Lab Testing for Introductory Cybersecurity

CHRISTIAN DELORENZO
SPRING 2023

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Cybersecurity Analysis and Operations
with honors in Cybersecurity Analysis and Operations

Reviewed and approved* by the following:

Nicklaus A. Giacobe, Ph.D.
Associate Teaching Professor of Information Sciences and Technology
Honors Adviser
Thesis Supervisor

Carleen Maitland, Ph.D.
Professor of Information Sciences and Technology
Faculty Reader

* Electronic approvals are on file.

ABSTRACT

This thesis discusses the current cybersecurity education landscape in academia and novel approaches to improve it. It goes over the shortcomings of the traditional cybersecurity education system and why there needs to be innovation in this space. The research aspect of this thesis focuses on the development and evaluation of a new version of PolyLab. PolyLab is a system of "game" levels designed to help teach the next generation of cybersecurity professionals the basics of using the Linux command line interface. It also has the added benefit of being polymorphic which prevents cheating and enables the user to have a unique experience with the system.

The proposed new version of PolyLab utilizes a browser-based virtual machine as the environment to run the "game" in. This approach enables students to have fast cost-effective access to the system from anywhere with an internet connection. This eliminates the need for expensive software or hardware and provides a safe environment for new students to practice their Linux skills without risking their system.

The research aspect of this thesis experimented on a group of underclassmen college students studying cybersecurity or related field. The study helped evaluate the new version of PolyLab and determine what systems to use to evaluate the system. Participants were asked to play the PolyLab "game" two times with time in between each session to measure their improvement and see how effective PolyLab is. Overall, this thesis provides valuable insights into the current state of cybersecurity education in academia and provides a good baseline for others to continue research and development in this field.

CONTENTS

CHAPTER 1 INTRODUCTION **1**

Linux in Cybersecurity 3
Cybersecurity Education 4
Solution 5

CHAPTER 2 LITERATURE REVIEW **8**

Review of Current Cybersecurity Education Systems 8
Related Academic Research 12

CHAPTER 3 BUILDING POLYLAB **15**

Building The Levels 15
Building The Linux Distribution and Server 18

CHAPTER 4 HUMAN SUBJECTS EXPERIMENT **23**

Results 33
Summary of Findings 47

CHAPTER 5 DISCUSSION AND CONCLUSION **49**

Literature Review 49
Building PolyLab 50
Human Subjects Experiment 50
Conclusions 51
Future Work 52

APPENDIX **53**

Institutional Review Board Application 53
Institutional Review Board Approval Letter 90
Demographics Survey 92
PolyLab Game Levels 98
Day 1 Post Lab Questionnaire 103
Day 2 Post Lab Questionnaire 108

BIBLIOGRAPHY **114**

ACADEMIC VITA **116**

LIST OF FIGURES

Figure 1. Example of Command Line 2

Figure 2 Email Address Prompt 5

Figure 3 Example User Management Level 16

Figure 4 Nginx Welcome Screen 22

Figure 5 Example Level 26

Figure 6 Microsoft Form 27

Figure 7 Example Website From v86 Project 47

LIST OF TABLES

Table 1 NASA-TLX Questions 30

Table 2 SART Questions 32

Table 3 Demographics Data 35

Table 4 Lab 1 (Part 1) 38

Table 5 Lab 1 (Part 2) 39

Table 6 Lab 1 (Part 3) 39

Table 7 Lab 1 (Part 4) 40

Table 8 Lab 1 (Part 5) 41

Table 9 Lab 2 (Part 1) 43

Table 10 Lab 2 (Part 2) 44

Table 11 Lab 2 (Part 3) 45

Table 12 Lab 2 (Part 4) 46

Table 13 Lab 2 (Part 5) 46

ACKNOWLEDGEMENTS

I would like to thank my honors advisor, Dr. Nicklaus Giacobe, for helping me through the process of writing my honors thesis. I am grateful for the countless hours he dedicated to this project. His expertise in the field of cybersecurity education was invaluable in shaping the research questions and providing overall support for this project. His mentorship was crucial in helping me become a better researcher. I would like to thank him for bringing me on board for this project.

I would like to express my gratitude to my friend, Eric Talanca. Without him, I would have been completely overwhelmed with this project. His experience in cybersecurity was essential in providing meaningful insights and feedback on the development of PolyLab. He devoted an extended amount of time to helping me recruit and manage the logistics of the lab sessions which was a great help.

I would also like to thank all the participants in the study. Without them, this research would not have been possible.

Chapter 1

Introduction

Cybersecurity is a constantly evolving landscape that has been around since almost the beginning of the internet. Cybersecurity is the practice of protecting critical digital systems and sensitive information from attacks. This goes hand in hand with information technology (IT) security as they combat threats against networked systems and applications, whether those threats come from inside or outside of a given organization. Cybersecurity has a rich history dating back to the birth of the internet with the Defense Advanced Research Projects Agency's (DARPA) Advanced Research Projects Agency Network (ARPANET) in the 1970s (*Arpanet*, 2023). This was the precursor to the internet as we know it today.

The first computer virus was created by Bob Thomas, a programmer at BBN Technologies in 1971 and a year later the first computer worm was created (*A Brief History of Computer Viruses*, n.d.). The first major cybersecurity incident that gained major media attention was the Morris Worm in 1988 (*Morris Worm*, n.d.). It was created by Robert Tappan Morris, a graduate student at Cornell University, and it was designed to gauge the size of the internet (*Morris Worm*, n.d.). The worm infected thousands of computers and caused damage to many of these systems. This has inevitably led to cybersecurity playing an ever more important role in our internet lives. It has evolved to include a wide range of technologies and practices, including firewalls, antivirus software, intrusion detection systems, and more (*Morris Worm*, n.d.).

As more and more devices come online with the advent of the internet of things (IoT), cybersecurity will continue to be a key industry in modern society. It is one of the most in-

demand jobs with a 2.72 million job shortage globally according to the National Institute of Standards and Technology (*Cybersecurity Workforce Demand*, n.d.). These are generally high-paying jobs with great opportunities for incoming employees. With the rise in these high-paying jobs, cybersecurity education has become an issue. There are simply not enough qualified people for all these jobs. Due to this, cybersecurity education needs to provide the necessary support to current and future cybersecurity professionals. This includes knowledge about hardware, software, various policies, and legal frameworks. As part of a cybersecurity curriculum, students will need the necessary technical and soft skills to be effective in their future jobs. Some of these include programming and system administration. While many of the incoming cybersecurity students have had some experience with computers and graphical user interfaces (GUI) such as Microsoft Windows or Apple macOS. Many of the incoming students will have little experience with the command line. The command line is a text-only user interface for interacting with the computer system. It allows a user to enter commands and see the output in text form. The command line is typically used by system administrators, developers, and power users as the command line gives the user greater access and control over a system. Users can make changes that simply are not available in the GUI. Some systems, like servers, don't have a GUI and require the command line to perform any configuration as seen in Figure 1.

Figure 1. Example of Command Line

Linux in Cybersecurity

Linux is one of the most popular operating systems, yet many people are unfamiliar with it. It is an open-source operating system based on UNIX and is UNIX compatible. It has been released under the GNU General Public License (GPL) enabling anyone to modify and use it for free in perpetuity (*What Is Linux?*, n.d.). As an operating system, it helps and directly manages the hardware, including CPU, memory, storage, and allocates the resources for various applications (*What Is Linux?*, n.d.). Linux is a popular server operating system due to its stability, security, and flexibility. It is also commonly used in embedded systems, cloud computing, and data science applications (*What Is Linux?*, n.d.). Linux has endless customizability that enables it to be one of the most popular operating systems in the cybersecurity industry.

When in a security role many cybersecurity professionals will find Linux systems hosting their organization's precious data or even hosting their web server. It can also be found on many cybersecurity tools. One of the most popular cybersecurity Linux distributions is Debian-based Kali Linux. It is designed for penetration testing, digital forensics, and security auditing (*What Is Kali Linux?*, n.d.). It is actively developed and maintained by Offensive Security. This Linux distribution comes included with many pre-installed tools that are used by cybersecurity professionals. This includes tools for network scanning, vulnerability analysis, password cracking, and exploitation (*What Is Kali Linux?*, n.d.). As a tool commonly used in the cybersecurity industry, it is also very popular in higher education to help teach cybersecurity concepts.

Cybersecurity Education

A prerequisite for using many of the included tools in Kali Linux is an intimate knowledge of the Linux command line. Many popular bundled applications such as Metasploit are command-line only and require the user to understand the basics of command-line applications to use them. Due to this requirement, many introductory cybersecurity classes need to teach the Linux command line to ensure that their students are competitive in the job market. Currently, there is a lack of applicable easy-to-use hands-on training for introductory cybersecurity courses. Many are locked behind paid services that do not provide a realistic and reliable service to teach Linux fundamentals. One popular example is Practice Labs. The service provides educational institutions with virtual machines that can be loaded with a variety of operating systems and configurations (*Practice Labs | Hands-on Learning for Digital & It Skills*, n.d.). The main problem with this and other services is their scalability, stability, and cost. A common issue is that these systems scale very poorly. It is quite common for the servers to become overwhelmed when students all log in and try to use the virtual machines at the same time (*Practice Labs Is Rated "Poor,"* 2023). These systems also lack the stability that is expected for a paid service. The cost also becomes an issue as each virtual machine is specific to each class. With a cybersecurity degree, one will have to pay for many virtual machine licenses as they will be taking many classes requiring virtual machines.

Solution

To help solve many of these issues, this research along with previous similar research has been building what we call PolyLab. The PolyLab project is designed to build a command-line Linux environment with interactive levels to help teach students the basics of the Linux terminal. As part of the new system, PolyLab will have additional advantages for students over conventional labs. This includes anti-cheat measures and the ability to do the labs over again for additional practice. There has been a rise in the use of sites such as Chegg or CourseHero currently to find answers to labs and tests, and PolyLab will make it significantly more difficult for these sites to enable students to cheat.

This system relies on multiple inputs and other data to build unique levels for each user. When the user first starts the system, they will be prompted for their email address as seen in Figure 2.

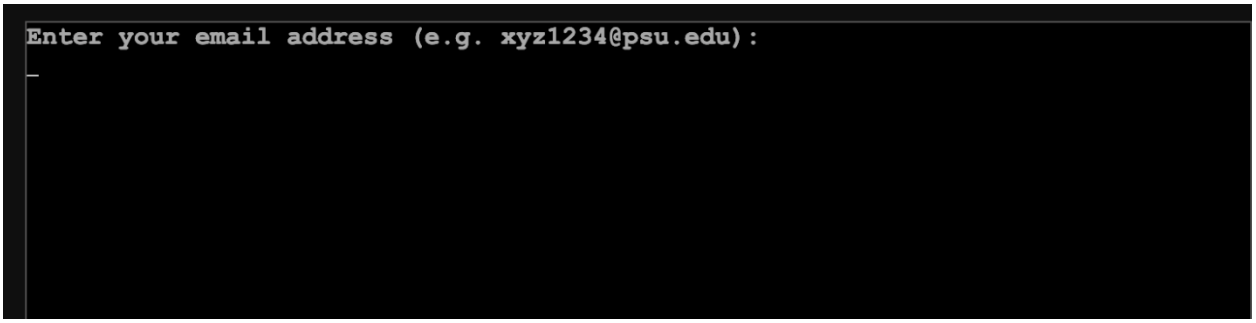


Figure 2 Email Address Prompt

The email address will be used along with the date and a password to build a hash which will be used for the answers and to customize the levels from word lists. Essentially a hash is a one-way fingerprint for electronic files. It is commonly used in cryptography, compression, checksum generation, and data indexing. One of the most popular uses for hashing is when a user enters a password on a website. Modern websites don't store your password as plaintext instead, they will

store a hash of your password. When the user logs into the website or service the system will compute the hash of the entered password on the fly and compare that to the stored hash of the password. PolyLab uses this function to build a pseudorandom seed for the application. This is a somewhat novel idea in cybersecurity teaching as most current labs have users following the same steps and can be easily shared between students or posted on homework-sharing websites. This implementation makes every environment unique for each user and even enables the user to come back the following day for a different set of answers.

On a technical level, another major improvement for this project was enabling it to be run in a browser completely on the front end. In this research study, we were able to iron out a lot of the issues with this implementation. It adds a lot of benefits to not only the end user but also for simply managing it. By having it run completely in the browser these students just learning will not need any additional software or licenses. The web server system will only need to host the static files and not need to do any processing for the system to operate. The core operating system that the PolyLab games run on top of is Buildroot. It is typically used in embedded systems as it is very small and efficient (*Buildroot - Making Embedded Linux Easy*, n.d.). This lab capitalizes on its small size to easily deliver it over the internet. All the processing of PolyLab is done by the user's browser with no input from the web server. This enables it to scale very easily eliminating one of the major issues with the competition which relies on compute from the host server to operate and can be easily overwhelmed.

To run in the browser, the lab virtualizes a custom Buildroot Linux distribution. It runs on top of v86 which is a project that emulates x86 CPUs and hardware. It translates the system calls into WebAssembly modules at runtime (copy, 2021). This provides adequate performance for the lab and runs better than the paid competition. V86 supports many operating systems

including Linux and Windows. This could potentially enable the researchers to improve the system and potentially enable more complex labs to be run in the browser without server-side input. The v86 system and the disk image iso file are hosted on an nginx webserver that the researchers have hosted on Microsoft's Azure cloud platform. This could easily be changed to the liking of future implementors. The researchers went with Ubuntu server for its relative ease of use and stability but could easily be implemented on a RHEL system on the Google Cloud Platform for example. As a fully open-source system, it provides a cost-effective and flexible system for its users and maintainers. Professors can easily build new levels and customize the system to fit their class environment.

Chapter 2

Literature Review

PolyLab is built on previous various literature and labs that enabled the system to be built. The literature review serves to identify the current state of cybersecurity education and understand how and where further research is needed to help future cybersecurity professionals. The first domain of this literature review will cover the current tools and technologies used in cybersecurity education, this will include environments, platforms, and even other games used by students. The information collected will be used to inform the development of new cybersecurity education programs and to identify areas where cybersecurity education can be enhanced.

Review of Current Cybersecurity Education Systems

One of the most popular systems in the cybersecurity education market is Practice Labs. It provides a platform for other vendors or universities to run virtualized learning environments for students to run and practice labs that were developed for the course. They provide their own labs as well as infrastructure for other organizations' labs as well. Practice Labs have a few benefits including that it is easy to maintain by the teaching staff as they are not responsible for the day-to-day upkeep of the supporting infrastructure and if they opt to use one of Practice Labs prebuilt labs then they also do not have to build any of the teaching materials (*Practice Lab*, n.d.). It also runs completely in the browser which makes it easy for the students as they do not need any additional software besides a web browser and an internet connection. This also is

much more secure for the user as beginner cybersecurity students might not understand all the configuration and security requirements for running and analyzing potential malicious data on their personal devices. It also easily integrates into learning management systems (*Practice Labs*, n.d.). A learning management system is a software application that primarily academic institutions use to administer, track, report, and deliver a course in part or completely over the internet. Common examples are Canvas and Blackboard. This allows students to submit and view assignments on semi-custom webpages for each course a student is enrolled in. This integration makes grading and delivering the content easier for both educators and students.

While Practice Labs does have a few advantages it also comes with many drawbacks. One of the major issues is instability. There have been near-constant reports of the service going down or crashing especially when a large number of users are hitting the same type of lab (*Practice Labs Is Rated "Poor,"* 2023). This could be mitigated if there were competing products in the same category, but this seems to not be the case. Nearly all other vendors seem to use Practice Labs under the hood. They also seem to have lackluster support which can be problematic when their systems are having problems (*Practice Labs Is Rated "Poor,"* 2023). This is especially problematic as this is a paid product. As these products are marketed to students this can put an additional financial strain on them while providing a suboptimal product. Another issue for some users is that their servers seem to be based in Europe. For students studying on other continents, this can increase latency as every interaction needs to be sent over undersea cables. Their client-server model does not help with this as all processing is done on their servers which can be very far away. The client computer is just a terminal for the user to access the remote VM. Overall, there should be a better solution available to students and faculty.

Another common active cybersecurity learning system is SEED labs. SEED labs is a hands-on lab exercise system to help educate incoming cybersecurity students. It was created by the SEED project with the support of the National Science Foundation (*Seed Project*, n.d.). According to SEED labs, it is used by over 1000 institutions (*Seed Project*, n.d.). It comes with a variety of advantages and positive aspects that enable it to be an effective teaching tool. As it is based on the 64-bit version of Ubuntu 20.04, it is also open source enabling anyone to modify it to meet their needs. This virtual machine comes free of charge which enables it to be accessible to all students. The author does have corresponding materials for sale but they are not necessary. It has practical and engaging labs for students to develop their critical thinking and problem-solving skills along with their technical knowledge. It also provides realistic security challenges and has important learning objectives for students.

While it is overall a good option for some, SEED labs does have some notable drawbacks. As a virtual machine (VM) image, it requires a hypervisor to be able to run (*Seed Project*, n.d.). A hypervisor is special computer software that enables other operating systems to be loaded onto a host. One physical machine can share its resources while still being segmented from the host enabling some security and ease of use. The main problem with using hypervisors is that they can consume a lot of resources and can be inefficient. This can be a major problem for lower-power systems that do not have sufficient memory and storage. VMs also require the user to maintain them which can be complicated for students just beginning to learn how to use Linux. The SEED labs system contains a lot of different learning objectives which is excessive for students just learning how to use the command line. Setting up a VM for the first time can also be a challenge for new students. Overall, SEED labs would be a good choice for more advanced students, but the lack of ease of use is a problem for new students.

Over the Wire Bandit game is another popular cybersecurity and Linux learning game. It offers many levels with increasing difficulty where the user must find the password to access the next level. Some of the concepts that are taught by Bandit include basic Linux commands, file permissions, scripting, encryption, and many more (*Overthewire: Bandit*, n.d.). To use the service, users require a Secure Shell (SSH) application. SSH is a network protocol that enables users to securely access remote servers over unsecured networks using public-key cryptography to authenticate. These additional software applications can be complex for students just learning how to use the command line as most SSH applications are command line only. Over the Wire generally does not provide enough explanation and helpful materials to new Linux users. This can be challenging to beginners who might not have enough background knowledge to help them through the levels.

Capture the flag (CTF) competitions are another great way for students to learn basic cybersecurity skills. CTF competitions ultimately have the goal of finding and retrieving a hidden piece of data, the flag, from a given system. There can be many formats for a CTF competition including attack-defense-style, jeopardy-style, and others (Harmon, 2016). Attack-defense-style CTFs typically have two teams, the attackers who attempt to break into the competition system and find the flag. The jeopardy-style CTF has participants solve problems or challenges to find the flag. CTFs can greatly help participants practice and learn cybersecurity skills like the Linux command line in a fun exciting environment (Harmon, 2016). It does come with a few drawbacks such as they are usually paid and have an entrance fee. They typically don’t make up for a higher education class as they are typically a timed event such as on the weekend. Many of these CTFs require some basic understanding of cybersecurity concepts as

they don't often have any hints or education sessions. Overall, they are a great way to supplement some other education systems.

Related Academic Research

Teaching cybersecurity effectively has been an ongoing issue in academia. The rapidly evolving nature of the cybersecurity industry makes it difficult for higher education to maintain awareness and continue to alter the curriculum to keep up with the industry. This particularly conflicts with the atmosphere at most colleges as they are very methodical, whereas cybersecurity and the larger technology industry have an attitude of moving fast and breaking things. Higher education institutions that teach cybersecurity lack many of the standards found in more traditional degree paths. Due to all of these issues, academia has already conducted some research into the matter.

This current research is built off previous implementations and investigations into polymorphic applications that teach core concepts in cybersecurity to students. Steve Cromity did a good job developing his new Linux command line instructional simulation called PolyBandit. It was based on the Over the Wire Bandit game but he added a polymorphic component to it (Steve J. Cromity III, 2021). This made the experience for each student unique which helped with learning and combatting cheating. His study focused on testing the differences between the control group and the polymorphic group. This between-groups experiment had one group use the static form of PolyBandit while another group uses the full polymorphic version of PolyBandit (Steve J. Cromity III, 2021). This research compared the differences between these groups when it comes to their engagement, retention, performance, and self-assessment. The

thesis concluded that the PolyBadit system is a scalable and helpful system for new Linux users which will help address problems and shortcomings in current cybersecurity education (Steve J. Cromity III, 2021). Steve Cromity's research was vital for this research as many of the levels that were built ended up in this study. This enabled the current study to explore new approaches to teaching core cybersecurity concepts. the current research study contributes to a broader body of knowledge on teaching core concepts in cybersecurity.

Polymorphic educational systems have been designed in other contexts. Ryan Kohler has developed his system for PolyLab that helped teach and prevent cheating in cybersecurity education. It primarily focused on the use of network data and other non-Linux cyber training (Ryan J. Kohler, 2018). During his research, he developed an automated grading system that was able to account for the pseudorandom answers that were created and unique to each user. The study also used a between-groups experiment to help determine if the polymorphic version provided added benefits to the user (Ryan J. Kohler, 2018). Ryan Kohler's thesis measures the situational awareness (SA) of the students, which is their ability to perceive, comprehend and project relevant information in the lab environment. The study found that participants did not report significantly higher workload or lower effectiveness than the students who received static network data showing that polymorphic applications can be an effective tool in cybersecurity education (Ryan J. Kohler, 2018). The study revealed that participants who received the polymorphic network data had higher engagement and interest in the assignment in comparison to the control this helps confirm this research's ability to help students.

SA plays an important role in understanding how effective PolyLab can be for students. Almost all previous PolyLab studies have relied upon Dr. Giacobe's research into SA in cybersecurity. This information has helped narrow down what questions to ask the participants

during the post-test questionnaire. The situation awareness rating technique (SART) and the perceived workload scale (NASA-TLX) were found to be effective for post-test analysis (Nicklaus A. Giacobe, 2013). This gives the researchers important insight into how a given instructional simulation affects the participant's cognitive load and ability to understand and respond to cybersecurity situations. The dissertation proposes a SA assessment battery for the cyber domain (Nicklaus A. Giacobe, 2013). These were adapted from other domains such as military command and control, aircraft piloting, and air traffic control, to suit the cyber-security context. This change is necessary because SA in the cyber domain is unique. Overall, this research is used to refine and improve PolyLab and other instructional tools in cybersecurity.

Chapter 3

Building PolyLab

The PolyLab project has two main components as part of the development and building process. The first component is building the levels and developing the learning experience while the other component is hosting and running the platform. The first component of the project demands a significant amount of creativity, as well as expertise in educational technology and game design. The second component involves the technical implementation of the project, including the setup of servers and supporting software.

Building The Levels

For this lab, the researchers utilized the previously built levels from PolyLab and integrate them with our ability to deliver the levels over a browser-based virtual machine. The levels that were used in this research are designed to teach and practice basic Linux commands. As part of this project and to help progress PolyLab in the future, additional levels were designed as part of this project. They were not tested in the lab environment. The primary focus of the additional levels was user management in a Linux environment.

User management is an essential part of Linux system administration. As a multi-user operating system, Linux supports multiple users working on the same system at the same time. Proper user management ensures that each user is given the appropriate level of access to the system and corresponding files, based on their role and responsibilities. For example, if a Linux system is configured as a storage server at an organization, the people in accounting do not need administrative privileges on the system. They also do not need access to the human resources

private records. This is an important concept to teach in cybersecurity curriculums as proper user management can prevent unauthorized access and prevent data breaches. User management is an important role in the CIA triad which stands for Confidentiality, Integrity, and Availability (*What Is the CIA Triad and Why Is It Important?*, n.d.). The CIA triad is a common cybersecurity guide and policy to protect sensitive data and ensure the reliability and availability of IT systems.

As part of the research into teaching Linux user management, example code was created that will be used in the future to expand on the original basic Linux command levels.

```
#!/bin/bash

# Read the user input

echo "Enter PSU ID: "
read ID
echo "The Current User ID is $ID"
hash="$(echo -n "$ID" | md5sum | cut -c1-32)"
echo "Create a new user called student. Then create a file called game.sh"
echo "Input this into the game.sh:"
echo "#!/bin/bash"
echo "echo password: qwerty12345"
echo
echo "Run this script again when done."


while read -r line; do
    user=$(echo "$line" | cut -d ':' -f 1)

        if [ "$user" = "student" ]; then
            uname2="$(stat --format '%U' "/home/student/game.sh")"
            if [ "${uname2}" = "student" ]; then
                echo "File was created by the student user. Now please make it executable. "
                if [ -x "/home/student/game.sh" ]; then
                    echo "Good job! Please submit the flag that is printed out by the game.sh file."
                else
                    echo "File is not an executable."
                fi
            else
                echo "It looks like the file was not created or not created by the correct user."
            fi
        fi
    done < "/etc/passwd"
```

Figure 3 Example User Management Level

As an example of a future level in the PolyLab system seen in Figure 3, this will test if a student knows how to create a new user, then change users so that they are the new user. It will then require them to know how to make a file using the touch command. The students will also need to know how to make the game.sh file executable so that they can get the password. In a full implementation of this level, the password that would be output on line 12 would implement polymorphism to help prevent cheating. This would be done by using the hash function on line 8 as the seed for the password that the user will use to confirm that the student was able to pass the level.

Other levels will include users learning to understand the /etc/passwd file and /etc/shadow file as they are key to user management. The /etc/passwd file contains information such as the user's username, user ID, group ID, GECOS, absolute path to the home directory, and login shell. The /etc/shadow file stores hashed and salted passwords that correspond to the users stored in /etc/passwd. The main problem with creating a polymorphic level with /etc/passwd is that most Linux distributions including Buildroot use salting when saving the hash in /etc/passwd. Hashing is the practice of adding a pseudorandom string of characters to the password before hashing it. This can prevent our authentication system from being able to predetermine the correct hash if a level were to be developed that asked a student to find the password hash of a given user.

The /etc/shadow file could relatively easily be used to build another level as it is somewhat predictable. In a future level, PolyLab would have multiple wordlists that would be used to help randomly generate some example users. The example usernames will have a small hash appended to the end, for example, "dog-abc123XYZ" which will be used to validate the correct answer in the submission form.

A final level could be built based on the /etc/group file. The /etc/group file stores information relating to which groups a user belongs to in a UNIX operating system. Each line has the group name, group ID (GID), and user list. Groups are important in Linux administration and cybersecurity as they provide a way to manage users with similar security and access privileges. This enables users of the same group to share files and other resources on the system. UNIX groups can also help prevent unauthorized access or modification of files by other users who are not members of the same group this makes it an important topic for cybersecurity students to learn. A PolyLab level for this concept would include a similar example as the /etc/shadow level. It will have randomly generated users and groups from wordlists. There will be a hash appended to the end like before to help validate the correct answer. To find the correct answer the user will need to find the group with a name that is different from the rest. As an example, there might be eight user groups and seven of them are names of animals while the last one is a name of a US city. All of the group names will have hashes appended to the end but only the different ones will have the correct hash.

Building The Linux Distribution and Server

In this implementation, we used David Humphrey's Browser VM GitHub project. It builds a custom Buildroot config for a Linux x86 VM. The custom Linux 4.15 kernel used removes unnecessary modules and drivers that are not needed on a browser-based virtual machine (Humphrey, 2021). The project builds a docker container that enables the creators to build all the config files and eventually build the Linux ISO that will be hosted on the webserver.

The first step is to clone the Browser VM GitHub project at https://github.com/humphd/browser-vm.git. The user will need to have docker and git already installed. Next, the container will need to be built by running “docker build -t buildroot .” The docker container will then need to be run by using the following commands:

```
docker run \
    --rm \
   --name build-v86 \
   -v $PWD/dist:/build \
  -v $PWD/buildroot-v86/:/buildroot-v86 \
   -ti \
  --entry point "bash" \
buildroot
```

This will run the build and drop you into a bash shell in the docker container. In the shell there will be directories including one called “board” inside this directory will be “v86.” The v86 directory will have another subdirectory called “rootfs_overlay.” This directory is used to override files in the root filesystem. For example, if the “etc” directory is added, then it will show up in the final output of the Linux ISO when it's finally built in the docker container. This is how the user can add additional levels to the base Linux distribution. The user will copy over the shell scripts and other required files that they built into the “rootfs_overlay” before building the ISO.

Next, the user will need to build the ISO after adding the files to “rootfs_overlay.” They will need to run “make BR2_EXTERNAL=/buildroot-v86 v86_defconfig” and then “make menuconfig,” This will drop the user into a text-based menu and the defaults should be used.

There will be another set of config files that need to be made by running the following commands “make savedefconfig” and “make linux-menuconfig.” Use the defaults again for the text-based menu that will show up. Finally, the user will need to run “make linux-savedefconfig” and then run “make.” This will take some time to run especially on slower computers. It is common to take a few hours for the Linux ISO to build. It is also recommended to save a snapshot at this stage if you are running this in a hypervisor. This will enable the user to not have to rebuild everything from scratch if they would like to add additional levels or modify the ISO in any way.

Once the Linux ISO is built the user will need to set up and create the host environment that the web server and a virtual machine will be served from. The ISO file will be named v86-linux.iso and will be dropped in the /dist directory. The file will need to be moved to the production system or if it’s the same system moved eventually to /var/www/. Our research system was hosted on a Microsoft Azure instance running Ubuntu 22.04 LTS. It can be run on almost any system that supports a web server, but Linux is recommended.

On the production system, v86 will need to be installed. This can be done by running running “wget https://github.com/copy/v86/archive/refs/tags/latest.tar.gz” in the home directory of the web server. Later the user can move over the files to the /var/www directory. The .gz file will need to be expanded by running “tar -xvzf latest.tar.gz.” This does not include three other files that enable the system to run the virtualized Linux distribution. They can be downloaded by running the following commands:

wget https://github.com/copy/v86/releases/download/latest/libv86.js

wget https://github.com/copy/v86/releases/download/latest/v86-fallback.wasm

wget https://github.com/copy/v86/releases/download/latest/v86.wasm

To get this to run the system will need the following dependencies installed:

```
sudo apt install make
sudo apt install openjdk-17-jdk build-essential libc6-i386 clang nasm gdb qemu-system
gcc rustfmt curl -y
sudo apt install nodejs
sudo curl https://sh.rustup.rs -sSf | sh
```

After these are installed, the user will need to configure rust by running:

```
rustup target add wasm32-unknown-unknown
source "$HOME/.cargo/env"
```

The code will then need to be compiled by entering the “v86-latest” directory and then running the following commands:

```
make
make all
```

Next, the user will need to install a webserver to serve the files to the users. In this research, we used nginx. It can be installed and started by running the following commands:

```
sudo apt install nginx -y
sudo systemctl enable nginx --now
```

To ensure that nginx is working the user can go to http://localhost and see a screen like in Figure 4.

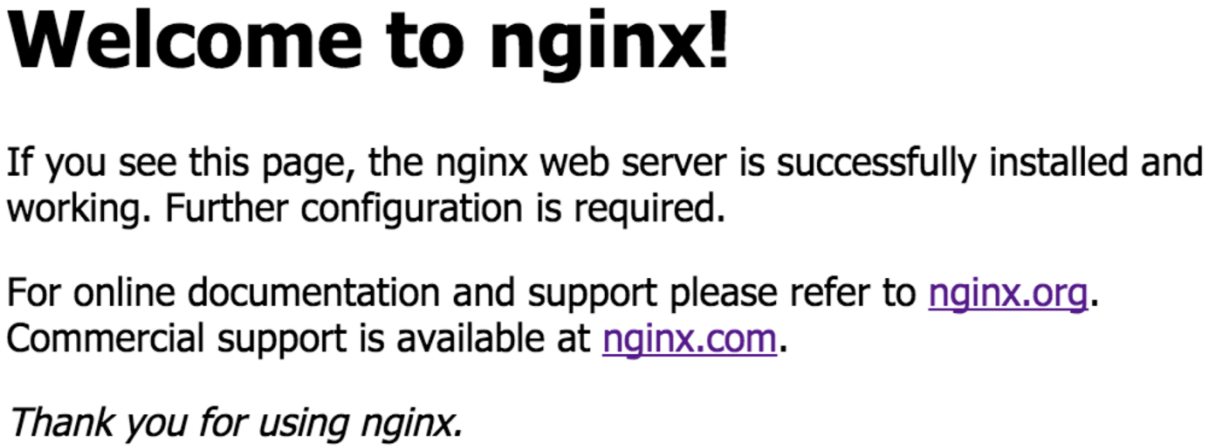


Figure 4 Nginx Welcome Screen

After nginx is confirmed to be working you can move all the JavaScript, web assembly, and HTML files to /var/www/. There is an examples directory included in v86 which can be used to easily get the system up and running. If you modify basic.html, which is found in the examples directory, the user can point the newly created image to basic.html enabling it to be run and accessible to the users. Here is an example of the change inside basic.html:

```
cdrom: {
    url: "../images/linux.iso",
},
```

The user can change this path to any file Linux ISO file that they might want to run and host on the web server.

Chapter 4

Human Subjects Experiment

The objective of this chapter is to describe and lay out the human subjects experiment that was run on study participants. This experiment studies the improvement that the participants go through between visits one and two using the polymorphic game levels to teach basic UNIX concepts. The experiment tries to answer the following research question:

Does using the games help improve engagement, retention, and performance, if any, between the first and second lab tests of a Linux instructional simulation wherein text files, directories, and user-facing messages change according to the user?

In this chapter, the design, implementation, and analysis of results for the human subjects experiment that was conducted as part of this thesis will be discussed. To help determine if there was an improvement in the teaching method using the polymorphic command line game, the experiment utilized two standard sets of questionaries, NASA-TLX and SART. The researchers hypothesize that users of the polymorphic version of PolyBandit will improve their retention which may be related to engagement and overall performance.

Background and Implementation

To help develop new teaching methods for introductory cybersecurity courses, this experiment will be used to help influence new labs for IST classes. Students will be introduced to basic Linux systems that will allow each student to learn the command line while not allowing

them to cheat on the lab. This will be done by creating an environment that gives each student a different version of the lab with different answers. Having this will ensure that each student must do the work themselves and cannot cheat from a fellow student. This is needed as students can use sites such as Chegg or CourseHero currently to find answers to labs and tests. This negatively impacts the student's learning from the course as they are not doing the work themselves. Having this polymorphic capability also enables students to play the games repeatedly while having new interactions with the game. Each time they play the game, they get a new environment with different variables. The levels are little mini-games that allow the user to interact with a UNIX command line. It enables the user to learn basic commands like changing directories and opening files. This all runs on an nginx webserver hosted on the Azure cloud platform.

As part of this preliminary research, we looked at having a maximum of 100 participants in the test, but we will use about 25 subjects for this small sample for this pilot study. This data could be used in future larger studies. The participants are primarily recruited from introductory courses in the College of IST and focus on underclassmen as this research focuses on updating the lower-level curriculum first. The research team will reach out to professors we know and ask them to mention our study in class. Professors can provide recruitment information in their courses but will not be answering questions about the study and rather providing study team member contact information. We provided them with a short description of the study:

The goal of the research is to create a lab environment for students to introduce them to basic Linux commands in a fun game-like environment. Students will be able to play the game multiple times with new variables every time. We see this as an advantage over the current system of Practice Labs and hope to augment and maybe replace it in the future. If you decide to

take part in the research, you will provide us with valuable information that will help the college of IST create better teaching aids and labs for future students. The information you have provided will be used in an honors thesis to help perfect new labs that will be used in IST 110 and Cyber 100. If you have any questions, please contact the researchers.

Experiment

The beginning of the study starts with each participant expressing interest in participating in the study by emailing or contacting an affiliated researcher. Once the potential participant reaches out to a researcher, the participant will be sent the demographics survey via email to ensure that they are eligible for the study. The survey will assess their demographics and skill level relating to the command line, with questions based on age, ethnicity, gender, and comfort with the UNIX terminal. They will also be provided a unique percipient ID which will be used for all future parts of the study.

If the participant is eligible for the study, then they will be directed to arrange a time for them and a researcher to meet over Zoom. They will be asked to share their screen. Participants are eligible if they are between 18 and 65 and have a computer with an internet connection. During the first meeting, the participants will be subjected to an introduction video on how the UNIX command line works to give them some background information. They will also be asked to confirm their continued consent to the study and willingness to proceed. Next, participants will be asked to open a web browser on their personal computer and navigate to a website hosting the study's application. The study’s website will allow the participant to access the features and functions of the application without having to download or install anything on their

computer. We will test that everything is working properly for all participants so that they will be able to complete the study without issue. We also provided them with instructions and guidance on how to use the application effectively.

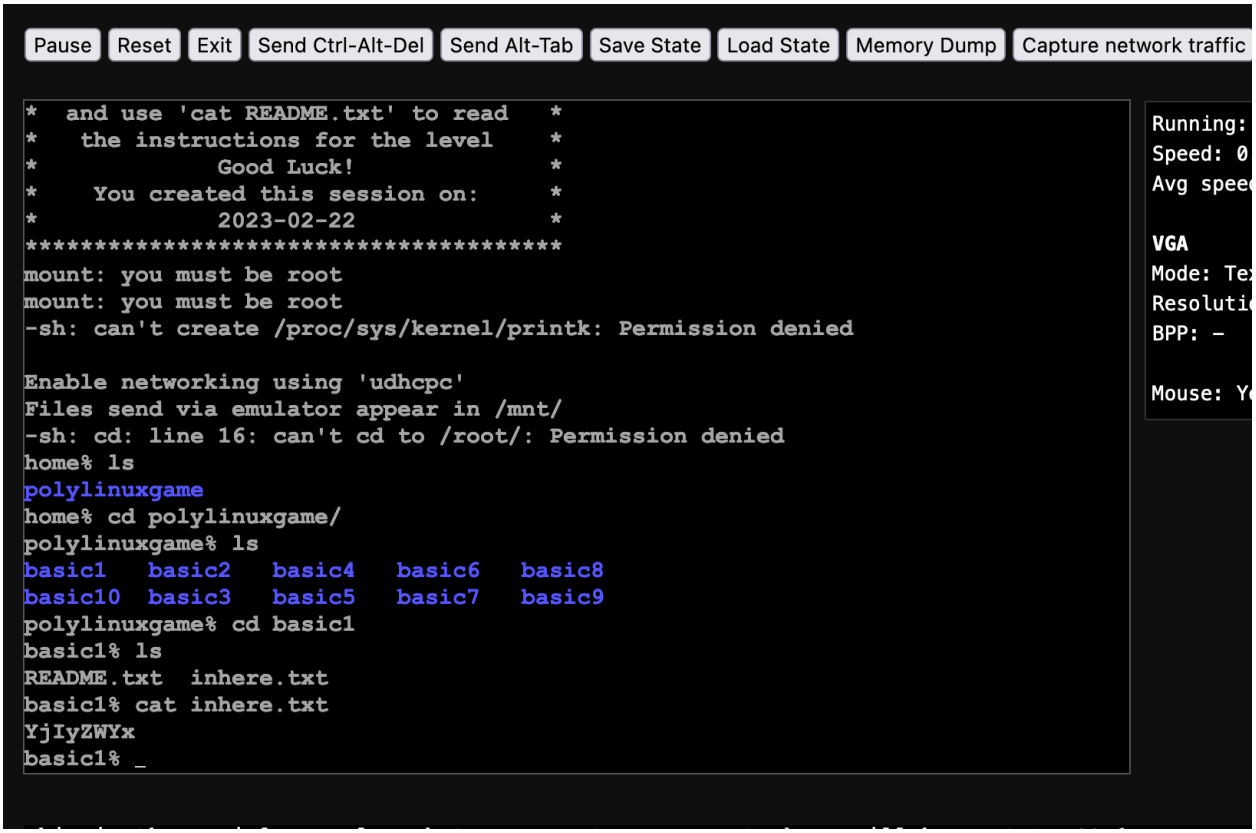


Figure 5 Example Level

Participants will be asked to start playing the levels of the game. An example level can be seen in Figure 5. Everyone will be started at the same time, and the Investigator will time everyone and record everyone's completion time along with their participant ID number which was given to them earlier in the study. The levels are little mini-games that allow the user to interact with and learn the UNIX command line. For example, it enables the user to learn basic commands like changing directories and opening files. The researchers designed the levels to be

fun and engaging, while also teaching the user essential skills for using a UNIX system. The polymorphic aspects of the levels enable the participants to play the games repeatedly while getting different answers and environments every time. The Investigator will monitor the participant's progress, performance, and collect data on how long they take to complete all 10 levels. The participant will submit the answers they find to a Microsoft form seen in Figure 6 which can check if the answer is correct and calculate their score based upon how many levels they got right.

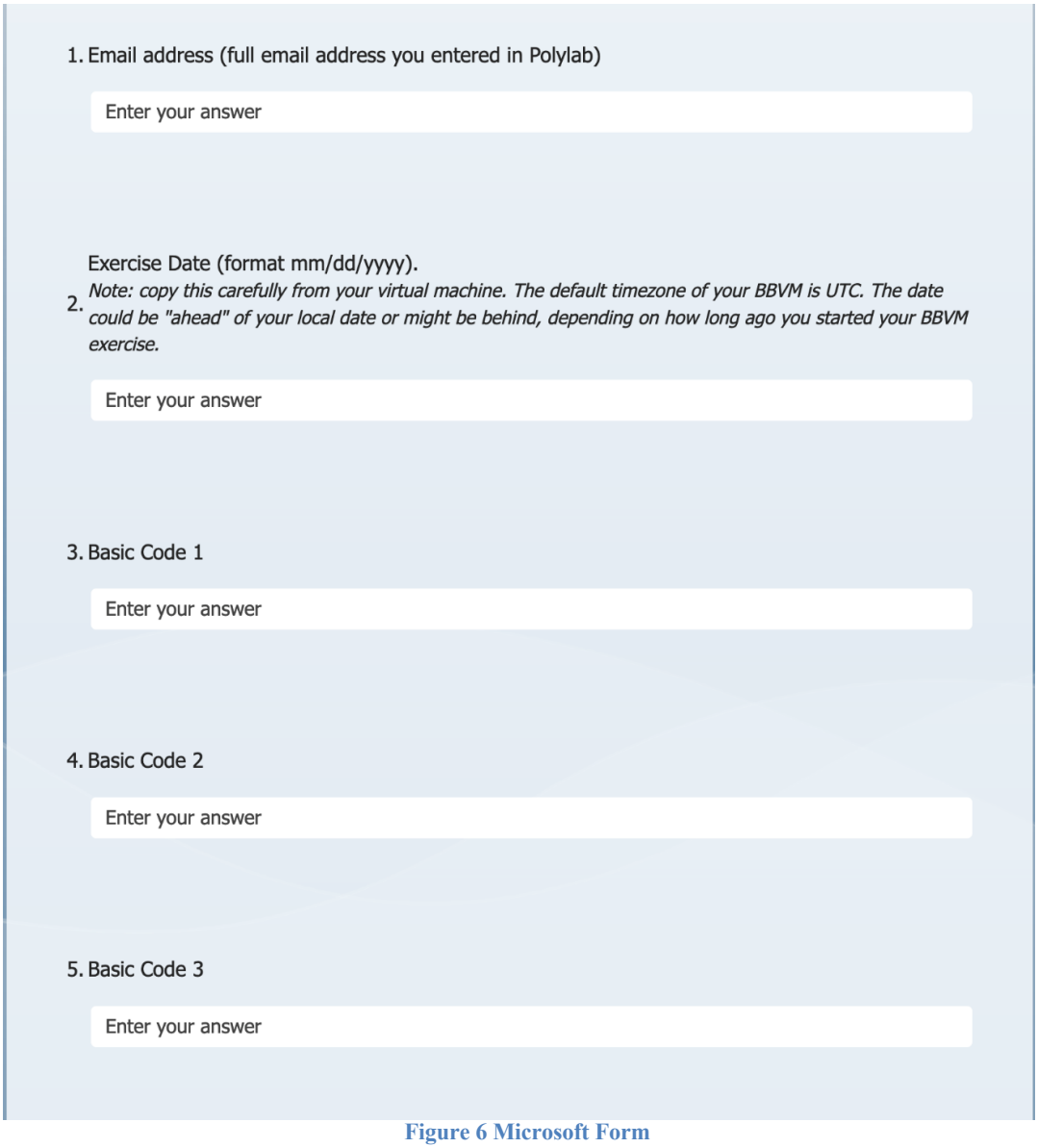


Figure 6 Microsoft Form

The form takes in the user's email and date so that it can build the expected correct answers. If a participant does not complete all the levels in 1.5 hours this will also be recorded by the Investigator and the testing will end. The participants will be asked to answer a short Google Forms questionnaire before leaving the Zoom room.

After the first session is completed, the subjects will be asked to schedule a subsequent visit to play the games again. Follow-up testing will be at least two weeks away but no longer than a month away. They will also be asked to continue to play the games outside of the testing environment. We will instruct the participants that playing the games will not be required, but we will ask them during the next visit how many times they played the game.

During the follow-up visit, participants will join the Investigator's Zoom meeting. The Investigator will ask and record how many times and for how long the participants played the games outside of the lab. The Investigator will ask all the Participants to navigate to the respective URL and get ready to play the levels again. Participants will play the levels again as they are polymorphic the levels will not be exactly the same but will require the same skills. They will have one hour to complete all the levels. After the participants are done, they will be asked to complete the second Google Form questionnaire.

Measures

In this lab, we have focused on collecting quantitative data from our participants through our use of Google Forms and the Investigators collecting time data and the maximum level the participant achieved if they were unable to complete all the levels in the specified time. This data is taken at the end of both lab visits. At the beginning of the first visit, the demographics survey data will be used to help correct for any bias. For example, as part of the demographics survey

we are asking the participants how much experience they have with Linux and the command line. If someone does not improve by any measurable amount and completes all the levels fairly quickly then the demographics survey will play an important role in determining if they already have experience with the Linux command line.

NASA-TLX

As part of the post-lab Google Form, we utilized the NASA-TLX tool that measures the workload a person experiences while using our system. This test looks at and measures mental demand, physical demand, temporal demand, performance, effort, and frustration. NASA-TLX was created in the 1980s by Sandra Hart a member of the Human Performance Group at NASA's Ames Research Center. It is seen as the gold standard for measuring subjective workloads across many different types of applications. The NASA-TLX questions have been rewritten to help the participant understand what we are looking for and direct their thinking to the specific context that we are researching. As part of our analysis, the data from the NASA-TLX questions will be used to help identify if the games were effective and enabled a better learning experience.

NASA-TLX Dimension	Description	Question Asked in Lab
Mental Demand	**how much thinking, deciding, or calculating was required to perform the task**	**How mentally demanding was the task in comparison to other computer related tasks?**
Physical Demand	**the amount and intensity of physical activity required to complete the task**	**How physically demanding was the task in comparison**

		to other computer related tasks?
Temporal Demand	**the amount of time pressure involved in completing the task**	**How hurried or rushed was the pace of the task in comparison to other computer related tasks?**
Performance	**how successful or satisfied the participant was with their performance on the task**	**How successful were you in accomplishing what you were asked to do?**
Effort	**how hard the participant had to work to achieve their level of performance**	**How hard did you have to work to accomplish your level of performance?**
Frustration	**how insecure, discouraged, irritated, stressed, or annoyed the participant felt during the task**	**How insecure, discouraged, irritated, stressed, and annoyed were you?**

Table 1 NASA-TLX Questions

SART

Situation Awareness Rating Technique (SART) is a subjective rating technique that was originally developed for the assessment of pilot situation awareness (SA). We use it in our post-trial phase to gain important insights into how effective the new teaching games are. SART was originally developed in 1990 for assessing pilot situation awareness but has been expanded to be used in many other areas. SART provides a subjective measure of situation awareness that reflects the operator's perception of their situation. This can be used to help identify the strengths and weaknesses of the operator's situation awareness and provide feedback that we can use to

help improve the games. SART also has some drawbacks including that relies on the operator's memory and judgment, which may be biased or inaccurate. This is still important information as someone's judgment is a key component of how students learn. It also can be influenced by the operator's mood, motivation, or confidence which could lead to varying results. SART questions are divided into different categories including instability, variability, complexity, arousal, mental capacity, concentration, attention, information quantity, information quality, and familiarity. As part of the study, the questions have been modified to help focus the participants, provide important context, and collect the data that the researchers need. As part of these changes, we removed the question on information quality as it was not needed. The participants are not experts in the field so they would not have a good understanding of quality in the context of the Unix command line.

SART Category	**Description**	**Question Asked in Lab**
Instability	**Likeliness of a situation to change suddenly**	**How UNSTABLE was the situation that was presented in the scenario?**
Variability	**Number of variables that require attention**	**How COMPLEX was the situation that was presented in the scenario?**
Complexity	**Degree of the complexity of the situation**	**How VARIABLE was the situation that was presented in the scenario?**
Arousal	**Degree that one is interested in the activity**	**How STIMULATING was the situation that was presented in the scenario?**

Mental Capacity	**Amount of mental ability available for new information**	**To what degree did the situation in the scenario require you to CONCENTRATE YOUR ATTENTION?**
Concentration	**Degree that one needs to use their thoughts and apply them to the given situation**	**To what degree did the situation in the scenario require you to DIVIDE YOUR ATTENTION?**
Attention	**Amount of division of attention in a given situation**	**While addressing the situation in the scenario, how much SPARE MENTAL CAPACITY would you say you had?**
Information Quantity	**Amount of knowledge received and understood by the user**	**Please rate the QUANTITY OF INFORMATION that was presented in the scenario.**
Familiarity	**Degree that the user is acquainted with the situation**	**How FAMILIAR does the situation in the scenario feel to you?**

Table 2 SART Questions

Results

Participants

The research team recruited twenty-four participants to be a part of the study. They primarily came from underclassmen studying cybersecurity or related field at a large public research university in the United States. Most of the recruited participants were from a computer networking class and most had limited experience with Linux and the command line. The recruited subjects were invited to two Zoom meetings over a week where they were able to play the levels and learn about the Linux command line. These Zoom meetings usually lasted 80 minutes for the first session and around 40 minutes for the second session. The primary age demographic was college students from 18-22 years old. 54% of the participant population self-identified as male while 37.5% identified as female. The majority of the participant population self-identified as Caucasian with their native language being English. This seems to line up with the demographics of the university where most of the students attend. The response rate on income was lower than the rest of the questions which is understandable as many people do not like talking about their finances. There was one participant who only answered the first three questions which can slightly skew the results as the population size is quite small. Overall, the demographics survey had an overall high response rate and provide needed context for the rest of our results.

Are you under the age of 18 or older than 65?	How long have you used the	How comfortable are you with the	How many IST or related	What year are you in school?

	Linux command line?	**UNIX terminal? (Scale from 1-5)**	**classes have you taken?**	
Yes: N=24 No: N=0	No experience: N=11 Less than 1 month: N=6 Less than 6 months but greater than 1 month: N=3 Less than 1 year but greater than 6 months: N=1 Less than 3 years but greater than 1 year: N=2 Greater than 3 years: N=1	1: N=13 2: N=6 3: N=3 4: N=1 5: N=1	14: N=1 8: N=1 7: N=1 6: N=1 5: N=4 4: N=3 3: N=8 2: N=2 1: N=1 Not answered: N=1	First year of college: N=8 Second year of college: N=4 Third year of college: N=6 Fourth year of college: N=5 Not answered: N=1
How old are you?	**What is your gender?**	**What is your ethnicity?**	**What is your native language?**	**What is your annual income?**
18: N=3 19: N=7 20: N=3 21: N=8 22: N=2	Male: N=13 Female: N=9	Arab: N=1 Asian: N=1	Arabic: N=1 English: N=19	Between $50,000-$80,000: N=1

Not answered: N=1	Rather not say: N=1 Not answered: N=1	Caucasian: N=16 From multiple races: N=1 Latino or Hispanic: N=2 Prefer not to say: N=2 Not answered: N=1	Spanish: N=2 Not answered: N=2	Less than $20,000: N=15 Prefer not to say: N=6 Not answered: N=2
What is your household income?	**What country are you from?**			
More than $125,000: N=5 Between $75,000-$125,000: N=1 Between $25,000-$75,000: N=5 Prefer not to say: N=10 Not answered: N=3	Dominican Republic: N=1 Mexico: N=1 United States: N=20 Not answered: N=2			

Table 3 Demographics Data

First Lab Testing

The first lab test had 22 total participants. Three people did not respond to the posttest and one person had technical difficulties with Zoom preventing them from completing the lab. The average time for the first lab was 37.43 minutes with an average number of correct answers is 5. There are a lot of variances with many participants getting nothing correct, while others get everything right. Upon further investigation into some submissions, certain students failed to put in the date correctly into the Microsoft Office form which was used to collect and grade the submissions. Due to the date being used as part of the hash that is used to build the levels, the form returns that the user got everything incorrect. Some participants also put in commands that they used to find the code instead of the actual code itself. This led them to get lower scores than what they should have gotten by inputting the code.

The first lab showed that the participants have a wide range of perceptions on how mentally demanding the lab was. Every level was represented by at least one participant with most (21.7%) selecting 4. The first lab was seen as not very physically demanding with over 65% of participants rating it as a 1 or 2. Most participants did not feel rushed through the tasks as 56.5% of them rated it at a 3 or lower. It was very interesting that most participants thought that they were very accomplished in the lab with 60.9% of participants rating it at 8 or above. This is in contrast with how hard the participants had to work to finish the lab, which was much more varied. The participants largely did not find the lab stressful or irritating with 67.2% of respondents rating it a 3 or under. The lab was also largely seen as stable with 43.5% of participants rating it a 1 and an additional 21.7% rating it a 2. The study was not seen as complex with 60.8% of participants rating it a 3 or under. It is hard to conclude how variable the lab was

as the respondents had a variety of opinions. The lab was seen as stimulating by 39.1% of the participants who rated it 8 or above. Participants had a variety of opinions on how much they needed to concentrate and divide their attention on the task. This was probably due to the varying skill levels participants came into the lab with. Participants' self-reported spare mental capacity was extremely varied due to their varied experience levels. The quantity of information being presented in the lab was centered around level 6. The familiarity of the information presented was divided with no participant picking the middle (5) this was largely due to how much experience they came into the lab with.

Participant ID	How mentally demanding was the task in comparison to other computer related tasks?	How physically demanding was the task in comparison to other computer related tasks?	How hurried or rushed was the pace of the task in comparison to other computer related tasks?
1002	2	2	1
1003	1	1	1
1004	Did not answer	Did not answer	Did not answer
1005	7	7	6
1006	10	1	5
1007	8	4	7
1008	2	1	2
1009	5	1	3
1010	7	1	3
1011	4	2	3
1012	4	1	3
1013	4	2	2
1014	2	2	3

1015	7	2	2
1016	8	3	7
1017	3	4	6
1018	Did not answer	Did not answer	Did not answer
1019	Did not answer	Did not answer	Did not answer
1020	4	3	7
1021	6	4	2
1022	4	2	2
1023	1	1	2
1024	5	1	7
1025	9	6	10

Table 4 Lab 1 (Part 1)

How successful were you in accomplishing what you were asked to do?	How hard did you have to work to accomplish your level of performance?	How insecure, discouraged, irritated, stressed, and annoyed were you?	How UNSTABLE was the situation that was presented in the scenario?
10	1	1	1
10	1	1	1
Did not answer	Did not answer	Did not answer	Did not answer
5	6	6	7
1	7	10	5
7	9	9	7
10	3	2	2
7	4	3	2
5	5	7	2
10	4	1	5
10	3	2	1
9	4	2	1
9	3	1	1
3	6	6	4
5	6	6	1
9	4	1	1
Did not answer	Did not answer	Did not answer	Did not answer
Did not answer	Did not answer	Did not answer	Did not answer
9	3	1	1
8	9	3	1
10	6	2	2

9	2	1	2
9	6	6	3
1	8	10	3

Table 5 Lab 1 (Part 2)

How COMPLEX was the situation that was presented in the scenario?	How VARIABLE was the situation that was presented in the scenario?	How STIMULATING was the situation that was presented in the scenario?	To what degree did the situation in the scenario require you to CONCENTRATE YOUR ATTENTION?
1	1	10	2
1	1	1	1
Did not answer	Did not answer	Did not answer	Did not answer
6	5	10	7
8	7	10	10
7	7	9	10
2	2	3	3
4	3	5	8
2	3	7	6
2	3	8	9
2	1	3	6
2	2	5	6
2	4	5	5
7	5	6	8
7	3	10	10
2	2	7	6
Did not answer	Did not answer	Did not answer	Did not answer
Did not answer	Did not answer	Did not answer	Did not answer
3	1	7	6
6	5	1	10
3	3	6	6
1	4	9	5
6	7	8	8
9	5	10	10

Table 6 Lab 1 (Part 3)

To what degree did the situation in the scenario require you to DIVIDE YOUR ATTENTION?	While addressing the situation in the scenario, how much SPARE MENTAL	Please rate the QUANTITY OF INFORMATION that was presented in the scenario.	How FAMILIAR does the situation in the scenario feel to you?

	CAPACITY would you say you had?		
5	10	1	3
1	10	3	10
Did not answer	Did not answer	Did not answer	Did not answer
6	6	9	7
6	7	8	6
7	4	8	3
2	8	4	9
5	4	2	7
4	5	9	8
2	8	8	6
2	9	5	8
5	8	3	10
6	8	6	1
2	3	6	4
5	7	10	3
6	4	6	8
Did not answer	Did not answer	Did not answer	Did not answer
Did not answer	Did not answer	Did not answer	Did not answer
3	6	5	6
10	7	10	2
7	5	6	8
1	2	5	2
3	4	7	6
9	2	7	3

Table 7 Lab 1 (Part 4)

Time it took to finish the lab	Number of correct levels
33:42:00	10
5:40	10
13:44	9
49:44:00	Answer sheet crashed
26:44:00	0
1 hour	0
11:58	10
29:55:00	6

43:10:00	9
30:00:00	0
20:40	0
27:53:00	9
30:38:00	10
29:00:00	0
31:37:00	3
20:25	0
technical difficulties and left	
17:57	10
13:54	9
36:00:00	9
29:29:00	9
37:55:00	0
34:20:00	9
37:25:00	1

Table 8 Lab 1 (Part 5)

Lab Two Testing:

The second lab test had 21 total participants. Three people did not show up to their follow-up lab. The average time for the second lab was 16.71 minutes with an average number of correct answers being 5.29. The time has shown significant improvement by more than halving. The correct answer average did not improve much as the time. This is likely due to similar issues with participants entering incorrect information into the date field which will mark their entries as incorrect when the system grades them.

During the posttest of the second lab, the participants were asked how many times they played the levels between their first and second tests. Almost all participants did not play the game in between the two different labs with 81% reporting 0 times. There was one outlier with 85 self-reported plays. There was a dramatic reduction in the mental demand with the second lab secession. 57% of participants reported a 3 or under when it comes to their mental demand. The

physical demand was similarly low with 83.44% of respondents reporting a 3 or under. No participants were rushed as no student reported anything higher than a 6. Participants felt very successful with 2/3 of the population reporting a 9 or above. There was still a variety of responses when it comes to how hard the participants had to work to accomplish their level of performance. Participants generally already did not feel insecure, discouraged, irritated, stressed, or annoyed and during the second lab, they felt even more so with 90.76% rating it a 3 or less. Similar to the previous question, the participants found that the lab was very stable with 90.4% of them reporting a 3 or less. Participants found the second lab less complex with no one reporting higher than a 7. The variability of the exercises remained largely reported under 5. The stimulation of the exercise seems to have gone down as the participants have gotten more familiar with the levels. Participants still had a variety of opinions on how much they needed to concentrate, but there was a change in their division of attention. Most participants reported their division of attention to be 5 or below. The participants space mental capacity was extremely variable ranging from a 1 to a 10. The quantity of information seems to have changed with most reporting 6 or less but with nearly a quarter reporting 9 or above. The participants were very familiar with the situation as they have already taken the lab before, and this shows in their 80.9% response of 8 or above.

Participant ID	How many times did you play the game between the first and second lab?	How mentally demanding was the task in comparison to other computer related tasks?	How physically demanding was the task in comparison to other computer related tasks?
1002	85	1	1
1003	0	1	1

1004	0	3	2
1005	0	7	1
1006	1	8	8
1007	0	5	3
1008	0	2	2
1009	0	4	1
1010	0	2	2
1011	0	3	3
1012	0	3	1
1013	0	1	1
1014	Did not answer	Did not answer	Did not answer
1015	2	2	1
1016	Did not answer	Did not answer	Did not answer
1017	1	4	2
1018	Did not answer	Did not answer	Did not answer
1019	0	2	2
1020	0	4	2
1021	0	6	5
1022	0	3	2
1023	0	1	1
1024	0	5	2
1025	0	7	7

Table 9 Lab 2 (Part 1)

How hurried or rushed was the pace of the task in comparison to other computer related tasks?	How successful were you in accomplishing what you were asked to do?	How hard did you have to work to accomplish your level of performance?	How insecure, discouraged, irritated, stressed, and annoyed were you?
1	10	10	1
1	10	1	1
2	10	4	1
5	10	7	1
6	3	6	7
4	9	6	2
2	10	2	2
2	9	4	2
1	7	5	3

3	10	3	1
3	10	3	1
1	10	1	1
Did not answer	Did not answer	Did not answer	Did not answer
1	5	3	2
Did not answer	Did not answer	Did not answer	Did not answer
3	10	3	1
Did not answer	Did not answer	Did not answer	Did not answer
3	9	2	2
6	10	3	1
3	5	6	3
2	9	3	2
1	1	1	1
6	8	4	2
3	4	8	9

Table 10 Lab 2 (Part 2)

How UNSTABLE was the situation that was presented in the scenario?	How COMPLEX was the situation that was presented in the scenario?	How VARIABLE was the situation that was presented in the scenario?	How STIMULATING was the situation that was presented in the scenario?
1	1	1	10
1	1	1	1
1	4	2	4
1	4	3	10
1	6	7	10
3	4	4	7
2	2	3	4
1	2	5	1
2	2	4	7
3	2	2	8
1	2	2	5
1	1	3	3
Did not answer	Did not answer	Did not answer	Did not answer
1	6	2	3
Did not answer	Did not answer	Did not answer	Did not answer
1	1	3	4

Did not answer	Did not answer	Did not answer	Did not answer
1	2	1	5
2	3	3	5
5	7	4	5
2	3	4	4
1	1	1	3
2	5	5	6
5	5	3	6

Table 11 Lab 2 (Part 3)

To what degree did the situation in the scenario require you to CONCENTRATE YOUR ATTENTION?	To what degree did the situation in the scenario require you to DIVIDE YOUR ATTENTION?	While addressing the situation in the scenario, how much SPARE MENTAL CAPACITY would you say you had?	Please rate the QUANTITY OF INFORMATION that was presented in the scenario.
1	1	10	1
1	1	10	3
7	4	7	4
8	1	4	10
10	8	9	10
6	4	5	9
5	3	9	2
3	2	1	2
4	3	8	9
8	4	8	9
7	5	5	4
1	1	10	1
Did not answer	Did not answer	Did not answer	Did not answer
8	3	3	6
Did not answer	Did not answer	Did not answer	Did not answer
4	1	8	4
Did not answer	Did not answer	Did not answer	Did not answer
3	2	7	5
7	3	5	5
8	8	5	6
7	4	7	6
2	1	1	2
6	3	9	5

6	4	2	6

Table 12 Lab 2 (Part 4)

How FAMILIAR does the situation in the scenario feel to you?	Time it took to finish the lab	Number of correct levels
10	8:42	10
8	4:30	10
8	14:00	10
3	33:40:00	6
8	29:42:00	2
9	18:26	10
9	5:52	10
7	24:40:00	8
10	14:50	9
9	7:42	0
9	6:42	0
10	14:00	9
Did not answer	Did not show up	
3	26:30:00	2
Did not answer	Family Emergency could not come	
10	19:00	0
Did not answer	Did not show up	
10	14:10	0
9	8:32	10
4	13:15	5
10	11:07	9
10	13:15	9
9	16:40	0
9	18:00	1

Table 13 Lab 2 (Part 5)

Summary of Findings

Overall, this research is unable to conclude anything conclusive as the data proved that the system is not mature enough for wider use in the cybersecurity education system. The lab did not have enough participants and the data we did get showed some significant flaws in the proposed system. One of the major issues is that people are unpredictable and their inputs into any given system can vary wildly without extreme input validation and training. As stated earlier, some participants tended to use seemingly odd dates for the input on the Microsoft Form which was used to calculate their grades. Some examples include participants putting in their participant IDs, the date in parenthesis, "Feb 16, 2021", and even what appears to be their birthday. This is after repeated reminders of the current date during the lab and reminding them what to put in the "Exercise Date" field. The date "Feb 16, 2021" was very odd as that was nearly two years ago, and a few people submitted it as the date. After doing further analysis that date seemed to come from the lab website. As shown the in Figure 7, it seems to refer to the version of the example website the lab was using from the v86 project.

Figure 7 Example Website From v86 Project

Participants also failed to understand what they should input into the code fields. Some failed to enter the code and instead put in the command they used to get the code. The participants were reminded of what needed to be put in each field. These inaccurate inputs exposed a weakness that was unexpected as there were thorough repeated instructions that were not followed or misunderstood by some participants. Another major issue is that the Linux terminal is sometimes hard to read. Some letters look very similar for example “O” and “0” look almost identical. “I”, “1”, and “l” also look the same causing some students to get graded incorrectly. The results were manually checked, and the scores were updated for students that had confusing letters in their codes. Some other minor findings include that the game might need to be more stimulating as only around 40% of participants reported it to be that way. Participants also stated that the quantity of information might be too much for some as the participants' responses are centered at level 6. Overall additional research and development are needed before full production and usage can be achieved in an educational environment.

Chapter 5

Discussion and Conclusion

This chapter summarizes the main insights that emerged from the research conducted in this thesis. This came from surveys, quantitative data from participants, and research about previous studies. Each section will go over the corresponding section and highlight the conclusions made from them. It will also go into the findings and how they relate to the cybersecurity education industry. This is important because it indicates that the study's findings are not only relevant to the specific context of the study but also have broader implications for the field of cybersecurity education.

Literature Review

The literature review provided important context and background information that enabled this study to be built. Previous versions of PolyLab and PolyBandit have plaid a crucial role in helping progress this thesis. Understanding how others have attacked the problem of cybersecurity education in higher education can provide valuable insights and inform the development of effective strategies which were used in this research. Steve Cromity's research was invaluable in this aspect as this study was very much an extension of what he has already accomplished. Dr. Giacobe's work on SA provided the backbone of how this research was going to test its participants and ensure that the new version of PolyLab would be effective. It was the inspiration for using NASA-TLX and SART in the post-test questionnaire that participants were

asked to answer. Overall, the literature review played a crucial role in the development and implementation of this study. Drawing on these resources and insights, this research was able to contribute to PolyLab and cybersecurity education as a whole.

Building PolyLab

Chapter 3 focused on the design and implementation of the new version of PolyLab. It primarily goes into detail about implementing a custom browser-based virtual machine for students to learn the basic commands on a live Linux command line. These instructions show the tools and technologies needed and used to build a similar system for cybersecurity education. This includes the use of David Humphrey's browser-vm project to build the custom version of Buildroot Linux and the v86 project to run the virtual machine in the browser.

The browser-based virtual machine aimed to provide a simple interface that was already preconfigured for students to learn. Virtual machines also had the added benefit of providing a safe environment for learning that would prevent a student from inadvertently fouling up their personal computer. This chapter also provided examples of how levels were built to help support students in their learning and provide them with fun challenging levels for them to practice.

Human Subjects Experiment

Chapter 4 addressed the human subject's experiment. The participants were asked about their engagement, retention, and performance during the lab. The lab also collected quantitative data such as how long the participants took to complete the lab and how many correct codes they

entered. The primary focus was looking for any improvement in the participants between the first and second visits to the virtual lab over Zoom.

The results showed that the system is not mature enough for use in cybersecurity academia. Participants performed many unexpected things during the lab that helped expose some of the flaws in the system including input validation. The Linux terminal also had some confusing letters making “0” and “O” look almost identical. This led to some participants mixing up these letters when submitting leading to inadvertent wrong grades. Overall, the human subject experiment uncovered some issues with the PolyLab system, but it provided valuable insights for further development and improvement. This shows the importance of conducting human subject experiments in the cybersecurity education field.

Conclusions

Overall, this thesis provides additional information to the cybersecurity educational system, even though the conclusions could not be conclusive due to the small sample size of the participants. Even with this limitation, this research provided valuable information surrounding the new version of PolyLab. This includes a better understanding of the potential issues arising from unexpected inputs that participants provided. By now enabling PolyLab to run in a browser, it can easily be distributed to students increasing portability and saving costs for the students. Although the study encountered some unexpected issues during the human subject experiment, these findings provided important feedback for PolyLab which can be addressed in the future.

The participants in the study showed a significant decrease in the time it took them to complete the lab from the first session to the second lab session. This might indicate that the

participants can quickly apply the skills they learned in the first session. The decrease in time also might suggest that participants were getting more comfortable with the new system and that they are becoming more familiar with the interface and the commands they will need to succeed in the lab. While the small sample size of participants limits the generalizability of this study, it has provided some valuable feedback for future versions.

Future Work

There is a need for further efforts in this domain to ensure a positive learning experience for new students in cybersecurity. There needs to be further development into input validation on the grading form. Currently, Microsoft Forms does not support input validation so alternatives that work with the current backend system need to be researched. Problems with characters looking similar need to be fixed. One way of fixing this issue is by having multiple correct codes created on the backend that will account for all the various characters that look the same. The current system only teaches and evaluates basic Linux commands. There was some preliminary research into future levels including user management but this needs to be fully built out in the future. User management would be a valuable addition to PolyLab as it plays a critical role in cybersecurity. Finally, further larger studies are needed to validate the results of the proposed fixes to PolyLab to make it a truly better system for cybersecurity students.

Appendix

Institutional Review Board Application

This appendix provides details regarding the Institutional Review Board Protocol utilized in the research conducted for this thesis.

HRP-591 - Protocol for
Human Subject Research

Protocol Title:
Provide the full title of the study as listed in item 1 on the "Basic Information" page in CATS IRB (http://irb.psu.edu).
Polymorphic Lab Testing for Introductory Cybersecurity

Principal Investigator:
Name: Chris DeLorenzo
Department: College of Information Sciences and Technology
Telephone: 2022868307
E-mail Address: cqd5463@psu.edu

Version Date:
Provide a version date for this document. This date must be updated each time this document is submitted to the IRB office with revisions. DO NOT revise the version date in the footer of this document.
03/12/23
Clinicaltrials.gov Registration #:
Provide the registration number for this study, if applicable. See "HRP-103- Investigator Manual", under "ClinicalTrials.gov" for more information.
N/A

Important Instructions for Using This Protocol Template:
This template is provided to help investigators prepare a protocol that includes the necessary information needed by the IRB to determine whether a study meets all applicable criteria for approval.

GENERAL INSTRUCTIONS[1]:
Prior to completing this protocol, ensure that you are using the most recent version by verifying the protocol template version date in the footer of this document with the current version provided in the CATS IRB Library.
Do not change the protocol template version date located in the footer of this document.
Some of the items may not be applicable to all types of research. If an item is not applicable, please indicate as such or skip question(s) if indicated in any of the instructional text.
GRAY INSTRUCTIONAL BOXES: Type your protocol responses below the gray instructional boxes of guidance language. If the section or item is not applicable, indicate not applicable.
Do NOT delete the instructional boxes from the final version of the protocol.
The protocol should be written in lay language. Do NOT copy and paste grant proposal information into the protocol.
Add the completed protocol template to your study in CATS IRB (http://irb.psu.edu) on the "Basic Information" page.

CATS IRB LIBRARY:
Documents referenced in this protocol template (e.g., SOP's, Worksheets, Checklists, and Templates) can be accessed by clicking the Library link in CATS IRB (http://irb.psu.edu).

PROTOCOL REVISIONS:
When making revisions to this protocol as requested by the IRB, please follow the instructions outlined in the guides available in the Help Center in CATS IRB (http://irb.psu.edu) for using track changes.
Update the Version Date on page 1 each time this document is submitted to the IRB office with revisions.

If you need help…

All locations:
Human Research Protection Program
Office for Research Protections
101 Technology Center
University Park, PA 16802-7014
Phone: 814-865-1775
Fax: 814-863-8699
Email: irb-orp@psu.edu
https://www.research.psu.edu/irb

[1] This template satisfies AAHRPP elements 1.7.B, I.8.B, I-9, II.2. A, II.2.I, II.3.A, II.3.B, II.3.C-II.3.C.1, II.3.D-F, II.4.A, III.1.C-F, II.2.D

Table of Contents

1.0 Objectives
2.0 Background
3.0 Inclusion and Exclusion Criteria
4.0 Recruitment Methods
5.0 Consent Process and Documentation
6.0 HIPAA Research Authorization and/or Waiver or Alteration of Authorization
7.0 Study Design and Procedures
8.0 Number of Subjects and Statistical Plan
9.0 Data and Safety Monitoring Plan
10.0 Risks
11.0 Potential Benefits to Subjects and Others
12.0 Sharing Results with Subjects
13.0 Subject Payment and/or Travel Reimbursements
14.0 Economic Burden to Subjects
15.0 Resources Available
16.0 Other Approvals
17.0 Multi-Site Study
18.0 Adverse Event Reporting
19.0 Study Monitoring, Auditing, and Inspecting
20.0 Future Undetermined Research: Data and Specimen Banking
21.0 References
22.0 Confidentiality, Privacy and Data Management

Objectives

Study Objectives
Describe the purpose, specific aims, or objectives. State the hypotheses to be tested.

The research questions of PolyLab are as follows:
RQ1: What are the differences in engagement, retention, and performance, if any, between a static and polymorphic version of a Linux instructional simulation wherein text files, directories, and user-facing messages change according to the user?

H1: Users of the Polymorphic version of PolyBandit will improve their retention which may be related to engagement, and overall performance.

Primary Study Endpoints
State the primary endpoints to be measured in the study.

Research typically has a primary objective or endpoint. Additional objectives and endpoints are secondary. The endpoints (or outcomes), determined for each study subject, are the quantitative measurements required by the objectives. Measuring the selected endpoints is the goal of a trial (examples: response rate and survival).

We are going to measure retention on 25 people that are going to be included in this pilot study.

Secondary Study Endpoints
State the secondary endpoints to be measured in the study.

N/A

Background

Scientific Background and Gaps
Describe the scientific background and gaps in current knowledge.

For clinical research studies being conducted at Penn State Health/Penn State College of Medicine, and for other non-PSH locations as applicable, describe the treatment/procedure that is considered standard of care (i.e., indicate how patients would be treated in non-investigational setting); and if applicable, indicate if the study procedure is available to patient without taking part in the study.

The goal of the research to create a lab environment for students to introduce them to basic Linux commands that will allow each student to remember the lessons taught while not

allowing them to cheat on the lab. This will be done by creating an environment that gives each student a different version of the lab with different answers. Having this will ensure that each student has to do the work themselves and cannot cheat from a fellow student. This is needed as students are able to use sites such as Chegg or CourseHero currently to find answers to labs and tests. This negatively impacts the student's learning from the course as they are not doing the work themselves.

Previous Data
Describe any relevant preliminary data.

N/A
Study Rationale
Provide the scientific rationale for the research.

This study will use a control group that will study the topics of Cyber 100 as they were taught previously and compare them to the PolyLab group. These groups will be recruited from introductory cyber courses in the College of IST, and focus on underclassmen.

Inclusion and Exclusion Criteria
Create a numbered list below in sections 3.1 and 3.2 of criteria subjects must meet to be eligible for study enrollment (e.g., age, gender, diagnosis, etc.).

Vulnerable Populations:

You MAY NOT include members of these populations as subjects in your research unless you indicate this in your inclusion criteria because specific regulations apply to studies that involve vulnerable populations.
The checklists referenced below outline the determinations to be made by the IRB when reviewing research involving these populations. Review the checklists as these will help to inform your responses throughout the remainder of the protocol.

Children –Review "HRP-416- Checklist - Children"
Pregnant Women – Review "HRP-412- Checklist - Pregnant Women"
Cognitively Impaired Adults- Review "HRP-417- Checklist - Cognitively Impaired Adults"
Prisoners- Review "HRP-415- Checklist - Prisoners"
Neonates of uncertain viability or non-viable neonates- Review "HRP-413- Checklist - Non-Viable Neonates" or "HRP-414- Checklist - Neonates of Uncertain Viability"
[Do not type here]

Inclusion Criteria
Create a numbered list of the inclusion criteria that define who will be included in your final study sample (e.g., age, gender, condition, etc.)

Subjects will between at least 18 years of age and less than 65 years of age.
Subjects must have the ability to use a standard internet connected computer.

3.1.1 Does **this research involve collecting data from individuals residing outside of the US?**
☒ No
☐ Yes – identify the countries where data collection will take place
[Type protocol text here]

Exclusion Criteria
Create a numbered list of the exclusion criteria that define who will be excluded in your study.

Subjects will be excluded if they report physical or cognitive disabilities that are inconsistent with standard computer usage.

Early Withdrawal of Subjects

Criteria for removal from study
Insert subject withdrawal criteria (e.g., safety reasons, failure of subject to adhere to protocol requirements, subject consent withdrawal, disease progression, etc.).

[If subjects opt to be withdrawn from the study, their responses may be excluded from data analysis. We do not anticipate any other early withdrawal condition.

Follow-up for withdrawn subjects
Describe when and how to withdraw subjects from the study; the type and timing of the data to be collected for withdrawal of subjects; whether and how subjects are to be replaced; the follow-up for subjects withdrawn from investigational treatment.

No follow-up with withdrawn subjects will be attempted.

Recruitment Methods
Upload recruitment materials for your study in CATS IRB (http://irb.psu.edu). DO NOT include the actual recruitment wording in this protocol.
StudyFinder: If StudyFinder (http://studyfinder.psu.edu) is to be used for recruitment purposes, separate recruitment documents do not need to be uploaded in CATS IRB. The necessary information will be captured from the StudyFinder page in your CATS IRB study.
Any eligibility screening questions (verbal/phone scripts, email, etc.) used when contacting potential participants must be uploaded to your study in CATS IRB (http://irb.psu.edu).
[Do not type here]

Identification of subjects
Describe the source of subjects and the methods that will be used to identify potential subjects (e.g., organizational listservs, established recruitment databases, subject pools, medical or school records, interactions during a clinic visit, etc.). If not recruiting subjects directly (e.g., database query for eligible records or samples) state what will be queried, how and by whom.

StudyFinder:
If you intend to use StudyFinder (http://studyfinder.psu.edu) for recruitment purposes, include this method in this section.
Information provided in this protocol, including the description of study procedures, compensation, and recruitment, needs to be consistent with information provided on the StudyFinder page in your CATS IRB study.

For Penn State Health submissions using Enterprise Information Management (EIM) for recruitment, and for non-Hershey locations as applicable, attach your EIM Design Specification form in CATS IRB (http://irb.psu.edu). See "HRP-103- Investigator Manual, Study Recruitment" for additional information.

DO NOT include the actual recruitment material or wording in this protocol.

Subjects will be identified by advertising in computer and technology-related classes, club meetings, email, and flyers on IST bulletins. These groups will be provided with a short statement. We will contact the leaders (Professors or club heads) and they will pass it out to their members. These classes will include, but are not limited to Cyber 100, IST 110, IST 220, and SRA 221. Clubs that will be included in selection are Gamma Tau Phi of the College of IST.

Recruitment process
Describe how potential subjects first learn about this research opportunity or indicate 'not applicable' if subjects will not be prospectively recruited to participate in the research. Subject recruitment can involve various methods (e.g., approaching potential subjects in person, contacting potential subjects via email, letters, telephone, ResearchMatch, or advertising to a general public via flyers, websites, StudyFinder, newspaper, television, and radio etc.). DO NOT include the actual recruitment material or wording in this protocol.
[Do not type here]

How potential subjects will be recruited.

The subjects will be recruited via the classes they are taking. For example, professors in Cyber 100 might ask students to join the study. We will be focusing our efforts on lower-level IST and Cyber classes. The research team will reach out to Professors we know and ask them to mention our study in the class. Professors can provide recruitment information in their courses

but will not be answering questions about the study and rather providing study team member contact information. We will provide them with a short description of the study:

The goal of the research to create a lab environment for students to introduce them to basic Linux commands in a fun game-like environment. Students will be able to play the game multiple times with new variables every time. We see this as an advantage over the current system of Practice Labs and hope to augment and maybe replace it in the future. If you decide to take part in the research, you will provide us with invaluable information that will help the college of IST create better teaching aids and labs for future students. The information you have provided will be used in an honors thesis to help perfect new labs that will be used in IST 110 and Cyber 100. If you have any questions please contact the researchers.

Where potential subjects will be recruited.

They will be recruited in classes and IST clubs.

When potential subjects will be recruited.

Upon IRB review

Describe the eligibility screening process. Screening begins when the investigator obtains information about or from a prospective participant in order to determine their eligibility.

☐ Eligibility screening is occurring *before* consent* - describe the process below
☒ Eligibility screening is occurring *after* consent describe the process below
☐ Eligibility screening is occurring, consent is not being obtained in this research – describe the process below
☐ Not applicable - Eligibility screening is not being done in this research

We will be asking all participants if they are over the age of 18 and less then 65. We will also ask them if they are able to use an internet-based computer and if they have access to one. Screening questions occur as soon as consent occurs and that they will be removed from the study if they do not meet the criteria.

In some studies, these procedures may not take place unless HIPAA Authorization is obtained OR a waiver of HIPAA Authorization when applicable for the screening procedures is approved by the IRB.

*Unless informed consent is waived by the IRB, screening before consent is only permitted when screening activities are limited to the collection of information through oral or written

communication OR when identifiable private information or identifiable biospecimens is obtained by accessing records or stored identifiable biospecimens. Screening before consent is not permitted if data will be used for activities other than eligibility screening/recruitment (e.g., data analysis).

Consent Process and Documentation
Refer to the following materials:
The "HRP-090- SOP - Informed Consent Process for Research" outlines the process for obtaining informed consent.
The "HRP-091– SOP - Written Documentation of Consent" describes how the consent process will be documented.
The "HRP-314- Worksheet - Criteria for Approval" section 7 lists the required elements of consent.
The "HRP-312- Worksheet - Exemption Determination" includes information on requirements for the consent process for exempt research. In addition, the CATS IRB Library contains consent guidance and templates for exempt research.
The CATS IRB library contains various consent templates for expedited or full review research that are designed to include the required information.
Add the consent document(s) to your study in CATS IRB (http://irb.psu.edu). Links to Penn State's consent templates are available in the same location where they are uploaded. DO NOT include the actual consent wording in this protocol.
[Do not type here]

Consent Process:

Check all applicable boxes below:
☐ Informed consent will be sought and documented with a written consent form *[Complete Sections 5.2 and 5.6; If this is the only box checked, mark Sections 5.3, 5.4 and 5.5 as 'Not applicable']*

☒ Implied or verbal consent will be obtained – subjects will not sign a consent form (waiver of written documentation of consent) *[Complete Sections 5.2, 5.3 and 5.6; If this is the only box checked, mark Sections 5.4 and 5.5 as 'Not applicable']*

☐ Informed consent will be sought but some of the elements of informed consent will be omitted or altered (e.g., deception). *[Complete section 5.2, 5.4 and 5.6; If this is the only box checked, mark Section 5.5 as 'Not applicable']*

☐ Informed consent will not be obtained – request to completely waive the informed consent requirement. *[Complete Section 5.5; If this is the only box checked, mark Sections 5.2, 5.3, 5.4 and 5.6 as 'Not applicable']*

☒ Exempt Research: If you believe that the research activities outlined meet one or more of the criteria outlined in "HRP-312- Worksheet- Exemption Determination", check this box. By checking this box, you are verifying that the exempt consent process will disclose the following:

Penn State affiliation; name and contact information for the researcher and advisor (if the researcher is a student); the activities involve research; the procedures to be performed; participation is voluntary; that there are adequate provisions to maintain the privacy interests of subjects and the confidentiality of the data.

Note: If this box has been checked, mark Sections 5.3, 5.4, 5.5, and 5.6 as "Not applicable." If the investigator's assessment is inaccurate, an IRB Analyst will request revision to the protocol and ask that consent forms and recruitment materials be submitted. Except for exemptions where Limited IRB Review is required (see "HRP-312- Worksheet- Exemption Determination") or where otherwise requested by the IRB, consent forms and recruitment materials are generally not reviewed nor approved by the PSU HRPP for research undergoing exempt review.

Obtaining Informed Consent

Consent Process
Describe the consent process, including when and where it will take place.

We will be providing them with a demographics survey via Google Forms. There will be a consent question in the form. During each visit with a researcher, the participant will take our short questionnaire via Google Forms which will also ask for continuing consent. After a subject has expressed interest, the researchers will email the perspective participant a preview of a consent document. We will get official consent from the participants via our demographics survey which will have consent agreements at the beginning.

Coercion or Undue Influence during Consent
Describe the steps that will be taken to minimize the possibility of coercion or undue influence in the consent process.

Researchers will be instructed to explicitly say that participation in the study is completely voluntary and will allow the subject to withdraw at any time in the process. There is no compensation provided.

Waiver of Written Documentation of Consent
Review "HRP – 411 – Checklist – Waiver of Written Documentation of Consent."

Indicate which of the following conditions applies to this research:
☒ The research presents no more that minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context.

OR

☐ The only record linking the subject and the research would be the consent document and the principal risk would be potential harm resulting from a breach of confidentiality. Each subject will be asked whether the subject wants documentation linking the subject with the research, and the subject's wishes will govern. (*Note: This condition is not applicable for FDA-regulated research. If this category is chosen, include copies of a consent form and /or parental permission form for participants who want written documentation linking them to the research.)*

OR

☐ If the subjects or legally authorized representatives are members of a distinct cultural group or community in which signing forms is not the norm, that the research presents no more than minimal risk of harm to subjects and provided there is an appropriate alternative mechanism for documenting that informed consent was obtained. (*Note: This condition is not applicable for FDA-regulated research.)*

For distinct cultural groups, describe the alternative mechanism for documenting that informed consent was obtained:

[Type protocol text here]

Indicate what materials, if any, will be used to inform potential subjects about the research (e.g., a letter accompanying a questionnaire, verbal script, or implied consent form)

N/A

Informed consent will be sought but some of the elements of informed consent will be omitted or altered (e.g., deception).
Review "HRP-410-Checklist -Waiver or Alteration of Consent Process" to ensure that you have provided sufficient information.

Indicate the elements of informed consent to be omitted or altered

N/A

Indicate why the research could not practicably be carried out without the omission or alteration of consent elements

N/A

Describe why the research involves no more than minimal risk to subjects.

N/A

Describe why the alteration/omission will not adversely affect the rights and welfare of subjects.

N/A

If the research involves using identifiable private information or identifiable biospecimens, describe why the research could not practicably be carried out without using such information or biospecimens in an identifiable format.

N/A

Debriefing
Explain whether and how subjects will be debriefed after participation in the study. If subjects will not be debriefed, provide a justification for not doing so. Add any debriefing materials to the study in CATS IRB.
N/A

Informed consent will not be obtained – request to completely waive the informed consent requirement
Review “HRP-410-Checklist -Waiver or Alteration of Consent Process” to ensure that you have provided sufficient information.

Indicate why the research could not practicably be carried out without the waiver of consent

N/A
Describe why the research involves no more than minimal risk to subjects.

N/A
Describe why the alteration/omission will not adversely affect the rights and welfare of subjects.

N/A
If the research involves using identifiable private information or identifiable biospecimens, describe why the research could not practicably be carried out without using such information or biospecimens in an identifiable format.

N/A

Additional pertinent information after participation

Explain if subjects will be provided with additional pertinent information after participation.

N/A

Consent – Other Considerations

Non-English-Speaking Subjects

Indicate what language(s) other than English are understood by prospective subjects or representatives.

If subjects who do not speak English will be enrolled, describe the process to ensure that the oral and written information provided to those subjects will be in that language. Indicate the language that will be used by those obtaining consent.

Indicate whether the consent process will be documented in writing with the long form of the consent documentation or with the short form of the consent documentation. Review “HRP-091 –SOP- Written Documentation of Consent” and “HRP-103 -Investigator Manual” to ensure that you have provided sufficient information.

N/A

Cognitively Impaired Adults

Refer “HRP-417 -CHECKLIST- Cognitively Impaired Adults” for information about research involving cognitively impaired adults as subjects.

Capability of Providing Consent

Describe the process to determine whether an individual is capable of consent.

Given that the subject can understand the verbal consent and can give verbal consent, they will be considered cognitively capable for this study.

Adults Unable to Consent

Describe whether and how informed consent will be obtained from the legally authorized representative. Describe who will be allowed to provide informed consent. Describe the process used to determine these individual’s authority to consent to research.

For research conducted in the state of Pennsylvania, review “HRP-013 -SOP- Legally Authorized Representatives, Children and Guardians” to be aware of which individuals in the state of Pennsylvania meet the definition of “legally authorized representative.”

For research conducted outside of the state of Pennsylvania, provide information that describes which individuals are authorized under applicable law to consent on behalf of a prospective

subject to their participation in the procedure(s) involved in this research. One method of obtaining this information is to have a legal counsel or authority review your protocol along with the definition of “children” in “HRP-013 -SOP- Legally Authorized Representatives, Children, and Guardians.”

N/A
Assent of Adults Unable to Consent
Describe the process for assent of the subjects. Indicate whether assent will be required of all, some, or none of the subjects. If some, indicate which subjects will be required to assent and which will not.

If assent will not be obtained from some or all subjects, provide an explanation of why not.

Describe whether assent of the subjects will be documented and the process to document assent. The IRB allows the person obtaining assent to document assent on the consent document and does not routinely require assent documents and does not routinely require subjects to sign assent documents.

N/A

Subjects who are not yet adults (infants, children, teenagers)

Parental Permission
Describe whether and how parental permission will be obtained. If permission will be obtained from individuals other than parents, describe who will be allowed to provide permission. Describe the process used to determine these individual’s authority to consent to each child’s general medical care.

For research conducted in the state of Pennsylvania, review “HRP-013-SOP- Legally Authorized Representatives, Children and Guardians” to be aware of which individuals in the state of Pennsylvania meet the definition of “children.”

For research conducted outside of the state of Pennsylvania, provide information that describes which persons have not attained the legal age for consent to treatments or procedures involved in the research, under the applicable law of the jurisdiction in which research will be conducted. One method of obtaining this information is to have a legal counsel or authority review your protocol along with the definition of “children” in “HRP-013-SOP- Legally Authorized Representatives, Children, and Guardians.”

N/A
Assent of subjects who are not yet adults

Indicate whether assent will be obtained from all, some, or none of the children. If assent will be obtained from some children, indicate which children will be required to assent. When assent of children is obtained describe whether and how it will be documented.

N/A

HIPAA Research Authorization and/or Waiver or Alteration of Authorization
This section is about the access, use or disclosure of Protected Health Information (PHI). PHI is individually identifiable health information (i.e., health information containing one or more 18 identifiers) that is transmitted or maintained in any form or medium by a Covered Entity or its Business Associate. A Covered Entity is a health plan, a health care clearinghouse or health care provider who transmits health information in electronic form. See "HRP-103 -Investigator Manual" for a list of the 18 identifiers.

If requesting a waiver/alteration of HIPAA authorization, complete sections 6.2 and 6.3 in addition to section 6.1. The Privacy Rule permits waivers (or alterations) of authorization if the research meets certain conditions. Include only information that will be accessed with the waiver/alteration.
[Do not type here]

Authorization and/or Waiver or Alteration of Authorization for the Uses and Disclosures of PHI

Check all that apply:
☒ Not applicable, no identifiable protected health information (PHI) is accessed, used, or disclosed in this study. *[Mark all parts of sections 6.2 and 6.3 as not applicable]*

☐ Authorization will be obtained and documented as part of the consent process. *[If this is the only box checked, mark sections 6.2 and 6.3 as not applicable]*

☐ Partial waiver is requested for recruitment purposes only (Check this box if patients' medical records will be accessed to determine eligibility before consent/authorization has been obtained). *[Complete all parts of sections 6.2 and 6.3]*

☐ Full waiver is requested for entire research study (e.g., medical record review studies). *[Complete all parts of sections 6.2 and 6.3]*

☐ Alteration is requested to waive requirement for written documentation of authorization (verbal or implied authorization will be obtained). *[Complete all parts of sections 6.2 and 6.3]*

Waiver or Alteration of Authorization for the Uses and Disclosures of PHI

Access, use or disclosure of PHI representing no more than a minimal risk to the privacy of the individual

Plan to protect PHI from improper use or disclosure
Include the following statement as written – DO NOT ALTER OR DELETE unless this section is not applicable because the research does not involve a waiver of authorization. If the section is not applicable, remove the statement and indicate as not applicable.

N/A

Plan to destroy identifiers or a justification for retaining identifiers
Describe the plan to destroy the identifiers at the earliest opportunity consistent with the conduct of the research. Include when and how identifiers will be destroyed. If identifiers will be retained, provide the legal, health or research justification for retaining the identifiers.

N/A

Explanation for why the research could not practicably be conducted without access to and use of PHI
Provide reasons why this research could not practicably be carried out without access to and use of PHI.

N/A

Explanation for why the research could not practicably be conducted without the waiver or alteration of authorization
Provide reasons why this research could not practicably be carried out without the waiver or alternation of authorization.

N/A

Waiver or alteration of authorization statements of agreement
By submitting this study for review with a waiver of authorization, you agree to the following statement – DO NOT ALTER OR DELETE unless this section is not applicable because the research does not involve a waiver or alteration of authorization. If the section is not applicable, remove the statement and indicate as not applicable.

N/A

Study Design and Procedures
Data collection materials that will be seen or used by subjects in your study must be uploaded to CATS IRB (http://irb.psu.edu). DO NOT include any actual data collection materials in this protocol (e.g., actual survey or interview questions).

[Do not type here]

Study Design
Describe and explain the study design.

The design of the study is to determine if there are improvements in performance, engagement, and retention from the first test to the second test. Participants will play a game to help them learn the UNIX command line. We will track their improvements from when they first play it to the second time playing through the game.

Study Procedures
Provide a step-by-step description of all research procedures being conducted (broken down by visit, if applicable) including such information as below (where and when applicable); describe the following:
HOW: (e.g., data collection via interviews, focus groups, forms such as surveys and questionnaires, medical/school records, audio/video/digital recordings, photographs, EKG procedures, MRI, mobile devices such as electronic tablets/cell phones, observations, collection of specimens, experimental drug/device testing, manipulation of behavior/use of deception, computer games, etc.) For surveys, indicate if subjects are able to skip questions that they don't want to answer.

WHERE: (e.g., classrooms, labs, internet/online, places of business, medical settings, public spaces, etc.)

After individuals express interest in taking part in the experiment, they will schedule a time with the Principal Investigator to perform the testing. They will then receive a document with instructions on what they need to do before their participation.

Visit 1 or Day 1 or Pre-test, etc.
Provide a description of what procedures will be performed on visit 1 or day 1 or pre-test in order of how these will be done. If your study only involves one session or visit, use this section only and delete 7.2.2.

The researcher will create an identifier for the participant so we can track them throughout the study.
The participant will take a survey about their demographics and skill level relating to the command line. Some of the questions will be based on age, ethnicity, gender, and how comfortable they are with the UNIX terminal.
The participant will then be asked to open up a web browser of their choice on their personal computer and be directed to a website that will be hosting the application. We will test that everything is working properly for all participants. The application is run on a nginx webserver hosted on the Azure cloud platform.

Participants will then be subjected to a short introduction video to how the UNIX command line works and how user permissions can be managed from the command line.
Participants will then be asked to start taking the levels. Everyone will be started at the same time and the PI will time everyone and record everyone's completion time. The levels are little mini games that allow the user to interact with a UNIX command line. It enables the user to learn basic commands like changing directories and opening files.
The subjects will be given a notice on returning for a subsequent visit to play the same version they did before. They will also be asked to continue to play the games outside of the testing environment. We will instruct the participants that playing the games will not be required but we will ask them during the next visit how many times they played the game.
Have participants take the questionnaire on Google Forms.
Participants will be asked to schedule a follow-up testing time that is at least two weeks away but no longer than a month away.

Visit 2 or Day 2 or Post-test, etc. (If applicable)
Provide a description of what procedures will be performed on visit 2 or day 2 or post-test in order of how these will be done. If your study involves more than two sessions or visits replicate this section for each additional session or visit (e.g., 7.2.3, 7.2.4, etc.). If your study involves only one session or visit, delete this section.

Participants will join the PI's Zoom meeting.
Participants will be asked to estimate how much they played the games outside of the lab.
The PI will ask all the Participants to navigate to the respective URL and get ready to play the levels again. Participants will replay the same game that they played previously.
The PI will record how long each participant takes to complete the levels and how long it has been since they performed in the first lab.
Participants will be asked to fill out the the Google Form questionnaire for day two.

Duration of Participation
Describe how long subjects will be involved in this research study. Include the number of sessions and the duration of each session - consider the total number of minutes, hours, days, months, years, etc.

The participants will have approximately 1.5 hours for the first visit and 1 hour for the second. We will cut them off at this time if they have not completed the game. Participants will be asked to schedule a follow-up testing time that is at least two weeks away but no longer than a month away.

Number of Subjects and Statistical Plan

Number of Subjects

Indicate the maximum number of subjects to be accrued/enrolled, to include all persons who sign consent for the study. If applicable, distinguish between the number of subjects who are expected to be enrolled and screened, and the number of subjects needed to complete the research procedures (i.e., numbers of subjects excluding screen failures.)

We will have a maximum of 100 participants in the test.

Sample Size Determination
If applicable, provide a justification of the sample size outlined in section 8.1 to include reflections on, or calculations of, the power of the study.

We will use about 25 subjects for this small sample for this pilot study. This data could be used in future larger studies.
Statistical or Analytic Methods
Describe the statistical methods (or non-statistical methods of analysis) that will be employed.

T-test and ANOVA to examine within group differences.

Data and Safety Monitoring Plan
This section is required when research involves more than Minimal Risk to subjects as defined in "HRP-001 SOP- Definitions."

Minimal Risk is defined as the probability and magnitude of harm or discomfort anticipated in the research that are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests. For research involving prisoners, Minimal Risk is the probability and magnitude of physical or psychological harm that is normally encountered in the daily lives, or in the routine medical, dental, or psychological examination of healthy persons.

Please complete each section below if the research involves more than minimal risk to subjects or indicate not applicable.
[Do not type here]

Periodic evaluation of data
Describe the plan to periodically evaluate the data collected regarding both harms and benefits to determine whether subjects remain safe.

N/A
Data that are reviewed
Describe the data that are reviewed, including safety data, untoward events, and efficacy data.

N/A

Method of collection of safety information
Describe the method by which the safety information will be collected (e.g., with case report forms, at study visits, by telephone calls and with subjects).

N/A

Frequency of data collection
Describe the frequency of data collection, including when safety data collection starts.
N/A

Individuals reviewing the data
Identify the individuals who will review the data. The plan might include establishing a data and safety monitoring committee and a plan for reporting data monitoring committee findings to the IRB and the sponsor.

N/A
Frequency of review of cumulative data
Describe the frequency or periodicity of review of cumulative data.

N/A
Statistical tests
Describe the statistical tests for analyzing the safety data to determine whether harms are occurring.

N/A
Suspension of research
Describe any conditions that trigger an immediate suspension of research.

N/A
Risks
List the reasonably foreseeable risks, discomforts, hazards, or inconveniences to the subjects related the subjects' participation in the research. Include as may be useful for the IRB's consideration, a description of the probability, magnitude, duration, and reversibility of the risks. Consider all types of risk including physical, psychological, social, legal, and economic risks. Note: Loss of confidentiality is a potential risk when conducting human subject research and must be listed here.
If applicable, indicate which procedures may have risks to the subjects that are currently unforeseeable.
If applicable, indicate which procedures may have risks to an embryo or fetus should the subject be or become pregnant.
If applicable, describe risks to others who are not subjects.

There are extremely limited risks for this procedure. The participants will only have physical discomfort resulting from normal usage of a computer. If a participant is experiencing discomfort, they will be able to take breaks if needed. The risk of loss of confidentiality is limited as we are asking for limited information from the participants and no additional PII like SSN. Potentially there could be a loss of demographic data if our data gets leaked. All demographic data is optional and this should be a limiting factor. The data will be securely stored in file shares that will be only accessible to people running the study.

Potential Benefits to Subjects and Others

Potential Benefits to Subjects
Describe the potential benefits that individual subjects may experience from taking part in the research. If there is no direct benefit to subjects, indicate as such. Compensation is not considered a benefit. Compensation should be addressed in section 13.0.

The benefit for participants is that they will have the opportunity to learn about the UNIX command line which will help them in future classes and job opportunities.

Potential Benefits to Others
Describe the potential benefits to society or others.

In the future, this research will help introductory IST classes teach in a more engaging way that will hopefully be more applicable to the real world. This will help professors have better labs that will save participants money while providing a better learning experience to the participants. Hopefully this research will eventually be able to be adapted to higher level classes as well.

Sharing Results with Subjects
Describe whether results (study results or individual subject results, such as results of investigational diagnostic tests, genetic tests, or incidental findings) will be shared with subjects or others (e.g., the subject's primary care physicians) and if so, describe how information will be shared.

The results will not be shared with the participants, but they will be able to read the results in the honors thesis.

Subject Payment and/or Travel Reimbursements
Describe the amount, type (cash, check, gift card, other) and timing of any subject payment or travel reimbursement. If there is no subject payment or travel reimbursement, indicate as not applicable.

Extra or Course Credit: Describe the amount of credit and the available alternatives. Alternatives should be equal in time and effort to the amount of course or extra credit offered.

It is not acceptable to indicate that the amount of credit is to be determined or at the discretion of the instructor of the course.

Approved Subject Pool: Indicate which approved subject pool will be used; include in response below that course credit will be given and alternatives will be offered as per the approved subject pool procedures.

N/A

Economic Burden to Subjects

Costs
Describe any costs that subjects may be responsible for because of participation in the research.

N/A
Compensation for research-related injury
If the research involves more than Minimal Risk to subjects, describe the available compensation in the event of research related injury.

If there is no sponsor agreement that addresses compensation for medical care for research subjects with a research-related injury, include the following text as written - DO NOT ALTER OR DELETE:
It is the policy of the institution to provide neither financial compensation nor free medical treatment for research-related injury. In the event of injury resulting from this research, medical treatment is available but will be provided at the usual charge. Costs for the treatment of research-related injuries will be charged to subjects or their insurance carriers.

For sponsored research studies with a research agreement with the sponsor that addresses compensation for medical care for research-related injuries, include the following text as written - DO NOT ALTER OR DELETE:
It is the policy of the institution to provide neither financial compensation nor free medical treatment for research-related injury. In the event of injury resulting from this research, medical treatment is available but will be provided at the usual charge. Such charges may be paid by the study sponsor as outlined in the research agreement and explained in the consent form.

It is the policy of the institution to provide neither financial compensation nor free medical treatment for research-related injury. In the event of injury resulting from this research, medical treatment is available but will be provided at the usual charge. Costs for the treatment of research-related injuries will be charged to subjects or their insurance carriers.

Resources Available

Facilities and locations
Identify and describe the facilities, sites, and locations where recruitment and study procedures will be performed.

If research will be conducted outside the United States, describe site-specific regulations or customs affecting the research, and describe the process for obtaining local ethical review. Also, describe the principal investigator's experience conducting research at these locations and familiarity with local culture.

The recruitment will be conducted at Penn State during or after class. It can also happen during club meetings. The study procedures will be performed over Zoom. Participants will be able to work on the study from home or other places around campus as long as its quiet and not distracting.

Feasibility of recruiting the required number of subjects
Indicate the number of potential subjects to which the study team has access. Indicate the percentage of those potential subjects needed for recruitment.

There will be hundreds of potential participants available from all the different classes. We will only need a small portion if the available population.

PI Time devoted to conducting the research
Describe how the PI will ensure that a sufficient amount of time will be devoted to conducting and completing the research. Consider outside responsibilities as well as other on-going research for which the PI is responsible. Please only provide a response for the principal investigator – do not include information about any other study team members.

The PI is a full-time student at Penn State. He expects to spend about 10-15 hours per week working on the research.

Availability of medical or psychological resources
Describe the availability of medical or psychological resources that subjects might need as a result of their participation in the study.

N/A

Process for informing Study Team
Describe the training plans to ensure members of the research team are informed about the protocol and their duties.

We will continue our weekly meetings.
Other Approvals

Other Approvals from External Entities
Describe any approvals that will be obtained prior to commencing the research (e.g., from engaged cooperating institutions IRBs who are also reviewing the research and other required review committees, community leaders, schools, research locations where research is to be conducted by the Penn State investigator, funding agencies, etc.).

N/A
Internal PSU Ancillary Reviews
DO NOT ALTER OR DELETE:
Ancillary reviews are reviewed by other compliance groups or individuals within Penn State that inform the IRB's review of a new study or a modification to an existing study.

PSU IRB may set applicable ancillary reviews for your study. Please refer to the "HRP-309 Worksheet – Ancillary Review Matrix" for more information (found in the CATS Library).
[Do not type here]

Multi-Site Study
If this is a multi-site study (i.e., a study in which two or more institutions coordinate, with each institution completing all research activities outlined in a specific protocol) and the Penn State PI is the lead investigator, describe the processes to ensure communication among sites in the sections below.
[Do not type here]

Other sites
List the name and location of all other participating sites. Provide the name, qualifications and contact information for the principal investigator at each site and indicate which IRB will be reviewing the study at each site.

N/A

Communication Plans
Describe the plan for regular communication between the overall study director and the other sites to ensure that all sites have the most current version of the protocol, consent document, etc. Describe the process to ensure all modifications have been communicated to sites. Describe the process to ensure that all required approvals have been obtained at each site (including approval by the site's IRB of record). Describe the process for communication of problems with the research, interim results, and closure of the study.

N/A

Data Submission and Security Plan
Describe the process and schedule for data submission and provide the data security plan for data collected from other sites. Describe the process to ensure all engaged participating sites will safeguard data as required by local information security policies.

N/A

Subject Enrollment
Describe the procedures for coordination of subject enrollment and randomization for the overall project.

N/A
Reporting of Adverse Events and New Information
Describe how adverse events and other information will be reported from the clinical sites to the overall study director. Provide the timeframe for this reporting.
N/A
Audit and Monitoring Plans
Describe the process to ensure all local site investigators conduct the study appropriately. Describe any on-site auditing and monitoring plans for the study.

N/A
Adverse Event Reporting

Reporting Adverse Reactions and Unanticipated Problems to the Responsible IRB
By submitting this study for review, you agree to the following statement – DO NOT ALTER OR DELETE:

In accordance with applicable policies of The Pennsylvania State University Institutional Review Board (IRB), the investigator will report, to the IRB, any observed or reported harm (adverse event) experienced by a subject or other individual, which in the opinion of the investigator is determined to be (1) unexpected; and (2) probably related to the research procedures. Harms (adverse events) will be submitted to the IRB in accordance with the IRB policies and procedures.

Study Monitoring, Auditing, and Inspecting

Auditing and Inspecting
By submitting this study for review, you agree to the following statement – DO NOT ALTER OR DELETE:

The investigator will permit study-related monitoring, audits, and inspections by the Penn State quality assurance program office(s), IRB, the sponsor, and government regulatory bodies, of all study related documents (e.g., source documents, regulatory documents, data collection

instruments, study data etc.). The investigator will ensure the capability for inspections of applicable study-related facilities (e.g., pharmacy, diagnostic laboratory, etc.).

Future Undetermined Research: Data and Specimen Banking
If this study is collecting identifiable data and/or specimens that will be banked for future undetermined research, please describe this process in the sections below. This information should not conflict with information provided in section 22 below OR the "HRP-598 – Research Data Plan Review Form" regarding whether or not data and/or specimens will be associated with identifiers (directly or indirectly). If there are no plans to use identifiable data/specimens for future, undetermined research, then this section is NOT applicable.
[Do not type here]

Data and/or specimens being stored
Identify what data and/or specimens will be stored, and the data associated with each specimen.

N/A
Location of storage
Identify the location where the data and/or specimens will be stored.

N/A

Duration of storage
Identify how long the data and/or specimens will be stored. If data and/or specimens will be stored indefinitely, indicate such.
N/A
Access to data and/or specimens
Identify who will have access to the data and/or specimens.
N/A

Procedures to release data or specimens
Describe the procedures to release the data and/or specimens, including: the process to request a release, approvals required for release, who can obtain data and/or specimens, and the data to be provided with the specimens.

N/A
Process for returning results
Describe the process for returning results about the use of the data and/or specimens.

Password-protected files
References
List relevant references in the literature which highlight methods, controversies, and study outcomes.

Cromity, S. (n.d.). *Etd explore*. PolyBandit: Polymorphic Linux Command Line Simulatio; The Pennsylvania State University. Retrieved January 25, 2023, from https://etda.libraries.psu.edu/?search_field=all_fields&q=cromity
Giacobe, N., FitzPatrick, L., & Ruff, M. (n.d.). *Polybandit*. https://github.com/giacobe/PolyBandit
Home | linux journey. (n.d.). Retrieved January 25, 2023, from https://notes.siira.io/
Nick, G. (n.d.). *Polystego*.
PolyLab. (n.d.). Retrieved January 25, 2023, from http://polylab.ist.psu.edu/
V86. (n.d.). [Rust, C, Assembly, JavaScript, WebAssembly, HTML]. https://github.com/copy/v86

Confidentiality, Privacy and Data Management
IMPORTANT: The following section is required for all locations EXCEPT Penn State Health and the College of Medicine. Penn State Health and College of Medicine should skip this section and complete "HRP-598 Research Data Plan Review Form." In order to avoid redundancy, for this section state "See the Research Data Plan Review Form" if you are conducting Penn State Health research. Delete all other sub-sections of section 22.

For research being conducted at Penn State Health or by Penn State Health researchers only: The research data security and integrity plan is submitted using "HRP-598 – Research Data Plan Review Form."

In order to avoid redundancy, for this section state "See the Research Data Plan Review Form" if you are conducting Penn State Health research. Delete all sub-sections of section 22.

For all other research: Complete the following section. Please refer to PSU Policy AD95 for information regarding information classification and security standards and requirements. It is recommended that you work with local IT staff when planning to store, process, or access data electronically to ensure that your plan can be carried out locally and meets applicable requirements. If you have questions about Penn State's Policy AD95 or standards or need a consultation regarding data security, please contact Penn State IT – Information Security at security@psu.edu.

Which of the following identifiers will be recorded for the research project? Check all that apply. If none of the following identifiers will be recorded, do not check any of the boxes.

	Hard Copy Data	Electronic Stored Data
Names and/or initials (including on signed consent documents)	☐	☒
All geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes,	☐	☐
All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except	☐	☐

that such ages and elements may be aggregated into a single category of age 90 or older		
Telephone numbers	☐	☐
Fax numbers	☐	☐
Electronic mail addresses	☐	☒
Social security numbers	☐	☐
Medical record numbers	☐	☐
Health plan beneficiary numbers	☐	☐
Account numbers	☐	☐
Certificate/license numbers	☐	☐
Vehicle identifiers and serial numbers, including license plate numbers	☐	☐
Device identifiers and serial numbers	☐	☐
Web Universal Resource Locators (URLs)	☐	☐
Internet Protocol (IP) address numbers	☐	☐
Biometric identifiers, including finger and voice prints	☐	☐
Full face photographic images and any comparable images	☐	☐
Any other unique identifying number, characteristic, or code (such as the pathology number)	☐	☐
Study code number with linking list	☐	☐
Genomic sequence data	☐	☐
State ID numbers	☐	☐
Passport numbers	☐	☐
Driver's license numbers	☐	☐

If storing paper records of research data, answer the following questions:

Where will the paper records, including copies of signed consent forms, associated with this research study will be stored?

N/A

How will the paper records be secured?

N/A

How will access to the paper records be restricted to authorized project personnel?

N/A

If storing electronic records of research data, indicate where the electronic data associated with this research study will be stored. Check all that apply.

☐ Penn State-provided database application. Check which of the following database applications are being used (check all that apply):

☐ Penn State REDCap

☐ Other – Specify - provided and approved database application:
[Type protocol text here if box checked]

☐ Penn State, College, or Department IT file server
☐ Penn State OneDrive or SharePoint
☒ Penn State GoogleDrive
☐ Web-based system provided by the sponsor or cooperative group - Specify URL and contact information:
[Type protocol text here if box checked]

☐ Other – Specify the database application or server:
[Type protocol text here if box checked]

Provide details about the data security features or attach security documentation provided by sponsor or group:
[Type protocol text here if box checked]

Please visit **datastoragefinder.psu.edu** for assistance with identifying appropriate data storage options. If the software to be used does not appear on that site, a **software request form** must be completed.

If there is a list/key that links indirect identifiers (code numbers, participant IDs, etc.) to direct identifiers, that list must not be comingled (i.e., stored in the same location) as the identifiable data, including copies of signed informed consent forms. Additionally, access to that list/key must be restricted to authorized project personnel.

Is there a list/key that links code numbers to identifiers?
☐ Yes - explain how the list that links the code to identifiers is stored separately from coded data:
[Type protocol text here if box is checked]

☒ Not applicable, there is no list that links code numbers to identifiers. Skip to section 22.6.

Is there a list of people who have access to the list/key?
☐ Yes – explain how access to that list is restricted and why certain persons require access.
[Type protocol text here if box is checked]

☐ No – explain why not:
[Type protocol text here if box is checked]

Describe the mechanism in place to ensure only approved research personnel have access to the stored research data (electronic and paper).
☒ Password-protected files

☐ Role-based security
☐ Specify all other mechanisms used to ensure only permitted users have access to the stored research data:
[Type protocol text here if box is checked]

The use of mobile devices or wireless activity trackers to collect identifiable research data may have to be approved by Penn State IT - Information Security.

Will research data be collected and/or stored on a wireless activity tracker or mobile application or will the study team enter research data on a mobile device, such as an electronic tablet or cell phone?

☒ No – skip to 22.8
☐ Yes - answer the following questions:

Specify the provider of the tracker or mobile devices(s)/application
☐ Supplied by the sponsor
☐ Penn State owned device
☐ A personal device
☐ Other – Please specify source: [Type protocol text here if box is checked]

Specify the type(s) of tracker or mobile device(s)/application that will be used to capture data and all identifiers captured on the mobile device(s)/application. Please list all devices, and if more than one, the identifiers to be collected on each.

[Type protocol text here]

Specify the type of data collected on the tracker or mobile devices(s)/application.

[Type protocol text here]

Specify the application or website used to collect the data from the tracker or mobile device, if applicable.

[Type protocol text here]

Describe the measures taken to protect the confidentiality of the data collected on the tracker or mobile device(s)/application. Please address physical security of the device(s), electronic security, and secure transfer of data from device(s) to the previously indicated data/file storage location provided in section 22.3.

[Type protocol text here]
Specify the

The use of online survey tools and email to collect or send research data containing identifiers that represent more than minimal risk to subjects may have to be approved by Penn State IT - Information Security.

Will any research data be directly entered/sent by subjects over the internet or via email (e.g., data capture using on-line surveys/questionnaires, surveys via email, observation of chat rooms or blogs)?

☐ No – skip to 22.9
☒ Yes - answer the following questions:

Specify the identifiers collected over the internet or via email (Including IP addresses if IP addresses will be collected).

The participant's email will be collected. The IP address might be inadvertently collected by the web server and stored in a log file. This will not be used in the study. It will be used for debugging the system if we run into system issues and need to resolve network or accessibility issues.

Specify the type of data collected over the internet or via email.

Demographic information and study data will be collected via Google Forms questionnaire.

Describe the measures taken to protect the confidentiality of the data collected?

We will use a randomly generated identifier.

Describe how the research team will access the data once data collection is complete.

We will use spreadsheets and Google Drive.

If the research involves online surveys, list the name(s) of the service provider(s) that will be used for the survey(s) (e.g., REDCap, Penn State licensed Qualtrics, Survey Monkey, Zoomerang)? (Note: The IRB strongly recommends the use of REDCap for online surveys that obtain sensitive identifiable human subjects data.)

☐ Penn State REDCap
☐ Penn State Qualtrics
☐ Penn State Microsoft Forms
☒ Penn State Google Forms
☐ Other - Please specify:

Application: [Type protocol text here]
URL (If applicable): [Type protocol text here]

If the answer above is “Other” contact security@psu.edu for approval of an alternative data capture method

[Type protocol text here]

Depending on the nature of the subject matter involved, certain security requirements must be in place for the audio and/or video recording or photographing of subjects. If the subject matter presents more than minimal risk to the subjects, then, before completing the section below, please contact Penn State IT - Information Security at security@psu.edu to confirm whether these requirements are required.
Specify the
Will any type of recordings (e.g., audio or video) or photographs of the subjects be made during this study?
☒ No - skip to section 22.10
☐ Yes - answer the following questions:

What will be used to capture the audio/video/images? Give a brief description of content.
☐ Audio – Describe the intended content of the audio recording:
[Type protocol text here]

☐ Video – Describe the intended content of the video recording:
[Type protocol text here]

☐ Photographs of the subjects – Describe the intended content of the photographs:
[Type protocol text here]

☐ 3-D Images – Describe the intended content of the of 3-D images:
[Type protocol text here]

☐ Other - Specify:
[Type protocol text here]

How will the recordings/photographs/images be stored (electronically or physically)?

[Type protocol text here]

Where will the recordings/photographs/images be stored?

[Type protocol text here]

Who will have access to the recordings/photographs/images?

[Type protocol text here]

Will any of the recordings be transcribed?
☐ Not applicable
☐ No
☐ Yes – indicate who will be doing the transcribing?
[Type protocol text here]

Will the recordings/photographs be used for purposes other than this research study?
☐ No
☐ Yes - specify purpose(s) (e.g., publication, presentations, educational training, future undetermined research):
[Type protocol text here]
What type of r
Certificate of Confidentiality (COC) - Is the research biomedical, behavioral, clinical or other research that is funded by the National Institutes of Health (NIH)?
☐ Yes - check one of the following:
☐ The research involves human subjects as defined by the DHHS regulations (See Worksheet HRP-310).
☐ The research involves collecting or using biospecimens that are identifiable to an individual.
☐ If collecting or using biospecimens as part of the research, there is a small risk that some combination of the biospecimen, a request for the biospecimen, and other available data sources could be used to deduce the identity of an individual.
☐ The research involves the generation of individual level, human genomic data.

Note: If <u>any</u> of the 4 items above are checked, a COC is automatically issued by NIH and applies to the research. Information about the COC must be included in the consent form.

☒ No - answer the following question.
If the research is not funded by NIH, will the investigator apply for a COC for this research study?
☒ No
☐ Yes

Note: For research not funded by NIH, the IRB may require a COC if the research is collecting personally identifiable information and the information is sensitive and/or the research is collecting information that if disclosed could significantly harm or damage the subject.

What steps will be taken to protect subjects' privacy interests? (Check all that apply.)
☒ Identification and recruitment of potential subjects follows procedures consistent with privacy standards
☐ Consent discussion and research interventions will take place in a private setting
☒ Limiting the information being collected to only the minimum amount of data necessary to accomplish the research purposes

☒ Limiting the people with access to the identifiable research data to the minimum necessary as specified in the application and consent process
☐ Other – Specify:
[Type protocol text here]

What is the process for ensuring correctness of data entry?
☐ Double data entry to reduce risk of errors
☒ Electronic edit checks to ensure data being entered are not obviously incorrect
☒ Random internal quality and assurance checking of research data
☐ Direct entry by subjects
☐ Other - Specify:
[Type protocol text here]

Does this research involve the generation of large-scale human genomic data as defined in NIH Genomic Data Sharing Policy (http://gds.nih.gov)?
☒ No
☐ Yes – describe the plan for de-identifying the dataset before sharing it with NIH-designated data repositories.
[Type protocol text here]

Note: Data sharing with an NIH-designated data repository may require execution of an institutional certificate. Please review the 'Institutional Certification for NIH Genomic Data Sharing' section of the Investigator's Manual for information about seeking institutional certification.

Does this research involve data sharing to public/restricted data repositories or as part of a journal requirement?

Data sharing is an important part of rigorous scientific discovery and the validation of results. Planning for data sharing is *strongly recommended*.

Data sharing includes sharing of identifiable, coded, or de-identified data. The data can be shared with public or restricted data repositories. Increasingly, journals require the sharing of data as a stipulation for publication.
NIH-funded studies require data sharing, unless explicitly granted an exception from the NIH.

☒ Yes (may be required for publication and future grant submission)
☐ No

Does this research involve transfer or disclosure of data and/or specimens to and/or from Penn State?

☒ No - skip the remainder of section 22.15
☐ Yes - answer the following questions:

Check all that apply:
22.15.1 ☐ Data are being transferred or disclosed to Penn State
What is the name of the third party(ies) (the institution, sponsor, etc.) sending or providing the data?

[Type protocol text here]

Is the third party requiring us to sign a contract regarding the data?
22.15.1.1 ☐ Yes - this contract must go through the Office of Sponsored Programs
https://www.research.psu.edu/osp/overview-pages/data-use-agreements
22.15.1.2 ☐ No

22.15.2 ☐ Data are being transferred or disclosed from Penn State
What is the name(s) of the third party(ies) (the institution, sponsor, etc.) receiving or accessing the data?

[Type protocol text here]

Note: Data transfers or disclosures may require a Data Use Agreement (DUA).

22.15.3 ☐ Specimens are being transferred to Penn State
What is the name(s) of the third party(ies) (the institution, sponsor, etc.) sending the specimens?

[Type protocol text here]

22.15.4 ☐ Specimens are being transferred from Penn State
What is the name(s) of the third party(ies) (the institution, sponsor, etc.) receiving the specimens?

[Type protocol text here]

Note: All material transfers, either sending or receiving, require a Material Transfer Agreement (MTA). Please contact the Office of Technology Management for more information.

22.15.5 Describe how the data/specimens will be securely transferred or disclosed to/from the third party(ies).

[Type protocol text here]

22.15.6 How are the research data/specimens being transferred from and/or sent to the third party(ies)? Complete the appropriate section(s) and check all that apply within each completed section.

22.15.6.1 Data being transferred or disclosed to Penn State:

☐ Data are being received in aggregate/metrics (just counts, no individual data)

☐ De-identified individual data are being received and there is no linking list at either institution (no identifiers, or links to identifiers, such as code numbers)

☐ Coded research data without any identifiers are being received and the linking list remains with the entity sending the data; the recipient of the data will not have access to the linking list

☐ Coded research data with identifiers (such as dates and/or any of the identifiers listed in section 22.14.7 aside from Study Code) are being received and the linking list remains with the entity sending the data; the recipient of the data will not have access to the linking list

☐ Data with identifiers (such as dates and/or any of the identifiers listed in section 22.14.7) are being received and the linking list remains with the entity sending the data; the recipient of the data will have access to the linking list

☐ Data with identifiers along with the linking list are being received

☐ Other – Specify:

[Type protocol text here if box is checked]

22.15.6.2 Data being transferred or disclosed from Penn State:

☐ Data are being sent in aggregate/metrics (just counts, no individual data)

☐ De-identified individual data are being sent and there is no linking list at either institution (no identifiers, or links to identifiers, such as code numbers)

☐ Coded research data without any identifiers are being sent and the linking list remains with the entity sending the data; the recipient of the data will not have access to the linking list

☐ Coded research data with identifiers (such as dates and/or any of the identifiers listed in section 22.14.7 aside from Study Code) are being sent and the linking list remains with the entity sending the data; the recipient of the data will not have access to the linking list

☐ Data with identifiers (such as dates and/or any of the identifiers listed in section 22.14.7) are being sent and the linking list remains with the entity sending the data; the recipient of the data will have access to the linking list

☐ Data with identifiers along with the linking list are being sent

☐ Other – Specify:

[Type protocol text here if box is checked]

22.15.6.3 Specimens being transferred or disclosed to Penn State:

☐ De-identified specimens are being received and there is no linking list at either institution (no identifiers, or links to identifiers, such as code numbers)

☐ Coded specimens without any identifiers are being received and the linking list remains with the entity sending the specimens; the recipient of the specimens will not have access to the linking list

☐ Coded specimens with identifiers (such as dates and/or any of the identifiers listed in section 22.14.7 aside from Study Code) are being received and the linking list remains with the entity sending the specimens; the recipient of the specimens will not have access to the linking list

☐ Coded specimens with identifiers (such as dates and/or any of the identifiers listed in section 22.14.7) are being received and the linking list remains with the entity sending the specimens; the recipient of the specimens will have access to the linking list

☐ Coded specimens with identifiers along with the linking list are being received

☐ Other – Specify:

[Type protocol text here if box is checked]

22.15.6.4 Specimens being transferred or disclosed from Penn State:

☐ De-identified specimens are being sent and there is no linking list at either institution (no identifiers, or links to identifiers, such as code numbers)

☐ Coded specimens without any identifiers are being sent and the linking list remains with the entity sending the specimens; the recipient of the specimens will not have access to the linking list

☐ Coded specimens with identifiers (such as dates and/or any of the identifiers listed in section 22.14.7 aside from Study Code) are being sent and the linking list remains with the entity sending the specimens; the recipient of the specimens will not have access to the linking list

☐ Coded specimens with identifiers (such as dates and/or any of the identifiers listed in section 22.14.7) are being sent and the linking list remains with the entity sending the specimens; the recipient of the specimens will have access to the linking list

☐ Coded specimens with identifiers along with the linking list are being sent

☐ Other – Specify:

[Type protocol text here if box is checked]

22.15.7 If transferring data/specimens with identifiers to or from Penn State, which of the following identifiers will be included with the data/specimens? Check all that apply:

☐ Names	☐ Medical record numbers
☐ Initials	☐ Health plan beneficiary numbers
☐ Street address	☐ Account numbers
☐ City	☐ Certificate/license numbers
☐ Driver's License numbers	☐ Passport numbers
☐ State	☐ State ID numbers
☐ Zip Codes	☐ Vehicle identifiers and serial numbers, including license plate numbers
☐ County	☐ Device identifiers and serial numbers
☐ Geocodes	☐ Web Universal Resource Locators (URLs)
☐ Precincts	☐ Internet Protocol (IP) address numbers

☐ All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death	☐ Biometric identifiers, including finger and voice prints
☐ Ages > 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older	☐ Full face photographic images and any comparable images
☐ Telephone numbers	☐ Any other unique identifying number, characteristic, or code (such as the pathology number) Specify: [Type protocol text here if box is checked]
☐ Fax numbers	☐ Study code numbers
☐ Electronic mail addresses	☐ Master list linking study code numbers to subject(s)
☐ Social security numbers	☐ Genomic sequence data
	☐ Other – specify: [Type protocol text here if box is checked]

Institutional Review Board Approval Letter

The Institutional Review Board (IRB) Approval Letter is a document that approves the research study.

PennState

Office for Research Protections
Human Research Protection Program
Office of The Senior Vice President for Research
The Pennsylvania State University
101 Technology Center
University Park, PA 16802

814-865-1775
irb-orp@psu.edu
research.psu.edu/irb

EXEMPTION DETERMINATION

Date: March 14, 2023

From: Stephanie Flohr, PhD, IRB Analyst

To: Chris Delorenzo

Type of Submission:	Initial Study
Title of Study:	Polymorphic Lab Testing for Introductory Cybersecurity
Principal Investigator:	Chris Delorenzo
Study ID:	STUDY00022123
Submission ID:	STUDY00022123
Funding:	Not Applicable
Documents Approved:	• HRP-591 - Protocol for Human Subject Research IRB Updates v2.pdf (2.02), Category: IRB Protocol • Photos of Levels.docx (02/22/23), Category: Data Collection Instrument • Polylab Form Day 1.docx (03/13/23), Category: Data Collection Instrument • PolyLab Form Day 2.docx (03/13/23), Category: Data Collection Instrument • Polylab Initial Demographics Form v2.pdf (03/12/23), Category: Data Collection Instrument • video (02/22/23), Category: Other

The Office for Research Protections determined that the proposed activity, as described in the above-referenced submission, does not require formal IRB review because the research met the criteria for exempt research according to the policies of this institution and the provisions of applicable federal regulations.

Continuing Progress Reports are **not** required for exempt research. You must notify the IRB when the exempt research study is closed/completed by completing a continuing review in CATS IRB.

Changes to exempt research only need to be submitted to the Office for Research Protections in limited circumstances described in the below-referenced Investigator

PennState

Office for Research Protections
Human Research Protection Program
Office of The Senior Vice President for Research
The Pennsylvania State University
101 Technology Center
University Park, PA 16802

814-865-1775
irb-orp@psu.edu
research.psu.edu/irb

Manual. If changes are being considered and there are questions about whether IRB review is needed, please contact the Office for Research Protections.

Investigators are required to follow the requirements listed in the HRP-103 – Investigator Manual, which can be found by navigating to the IRB Library within CATS IRB (http://irb.psu.edu).

Investigators are also responsible for reviewing the History tab of their STUDY in CATS to ensure that any administrative HRPP requests are addressed in a timely manner.

This correspondence should be maintained with your records.

Demographics Survey

The following is the survey used to collect the demographics data on the participants.

Polylab Demographics Form

This form will help us gain important insight into the demographics of our participants and correct for any bias in our data. All the questions are optional besides the first four.

cqd5463@psu.edu (not shared) Switch account

* Required

Are you under the age of 18 or older then 65? *

- ◯ Yes
- ◯ No

Please enter your participant ID which the researcher gave you on your first visit. * Do not enter your PSU email.

Your answer

How long have you used the Linux command line?

- ◯ No experance
- ◯ Less than 1 month
- ◯ Less than 6 months but greater than 1 month
- ◯ Less than 1 year but greater than 6 months
- ◯ Less than 3 years but greater than 1 year
- ◯ Greater than 3 years

How comfortable are you with the UNIX terminal?

	1	2	3	4	5	
Not comfortable	○	○	○	○	○	Very comfortable

How many IST or related classes have you taken?

Your answer

What year are you in school?

- ○ First year of college
- ○ Second year of college
- ○ Third year of college
- ○ Fourth year of college
- ○ Fifth year of college
- ○ Greater than the fifth year of college
- ○ Not in college

How old are you?

Your answer

What is your gender?

- ○ Male
- ○ Female
- ○ Trans-girl – Someone who has or is currently transitioning from male to female.
- ○ Trans-boy – Someone who has or is currently transitioning from female to male.
- ○ Gender fluid – Those who have different gender identities at different times.
- ○ Agender – Those with no gender identity or a neutral identity.
- ○ Androgynous – Partly male and female. Not one specific sex.
- ○ Bi-gender – Those who experience two gender identities, either at the same time or swapping between the two. These can be male and female or other identities.
- ○ Non-binary – A blanket term to describe those who do not feel exclusively male or female.
- ○ Demi-boy – Someone whose identity is only partly male, regardless of their birth gender. They may or may not also identify as another gender.
- ○ Demi-girl – Someone whose identity is only partly female, regardless of their birth gender. They may or may not also identify as another gender.
- ○ Genderqueer – Those who don't go along with traditional gender distinctions.
- ○ Gender nonconforming – Those who do not follow conventional ideas about how they should look or act based on their birth gender.
- ○ All genders – Someone who identifies as every possible gender option.
- ○ Intersex – Someone with physical, genetic and hormonal features of a male and female.
- ○ Not sure
- ○ Rather not say
- ○ Other:

What is your ethnicity?

- ○ African-American
- ○ Asian
- ○ Caucasian
- ○ Latino or Hispanic
- ○ Native Hawaiian or Pacific Islander
- ○ From multiple races
- ○ Prefer not to say
- ○ Other:

What is your native language?

- ○ English
- ○ Spanish
- ○ French
- ○ Portuguese
- ○ Arabic
- ○ Mandarin
- ○ Cantonese
- ○ Hindi
- ○ Prefer not to say
- ○ Other:

What is your annual income?

- ◯ Less than $20,000
- ◯ Between $20,000-$50,000
- ◯ Between $50,000-$80,000
- ◯ More than $80,000
- ◯ Prefer not to say

What is your **household** income?

- ◯ Less than $25,000
- ◯ Between $25,000-$75,000
- ◯ Between $75,000-$125,000
- ◯ More than $125,000
- ◯ Prefer not to say

What country are you from?

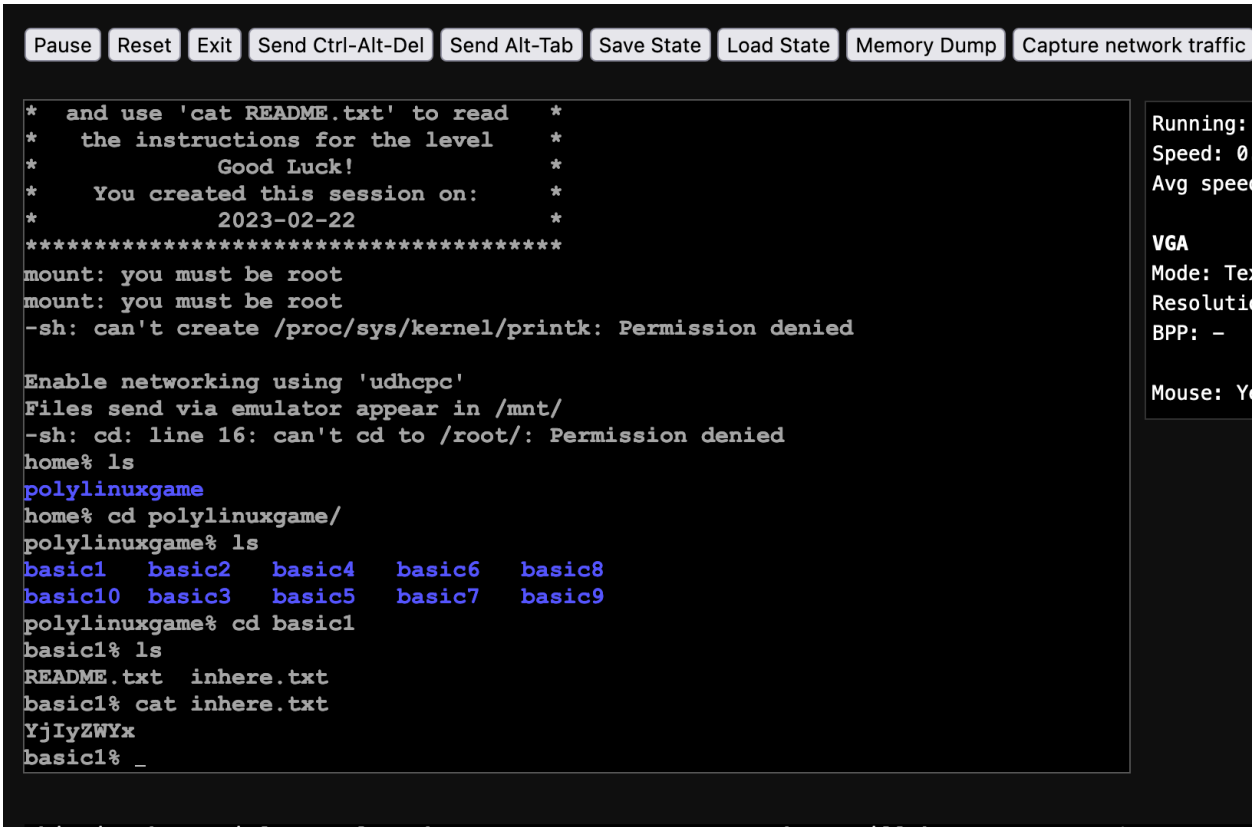
Choose

Submit Clear form

Never submit passwords through Google Forms.

PolyLab Game Levels

The following appendix is all the game levels the participants played.

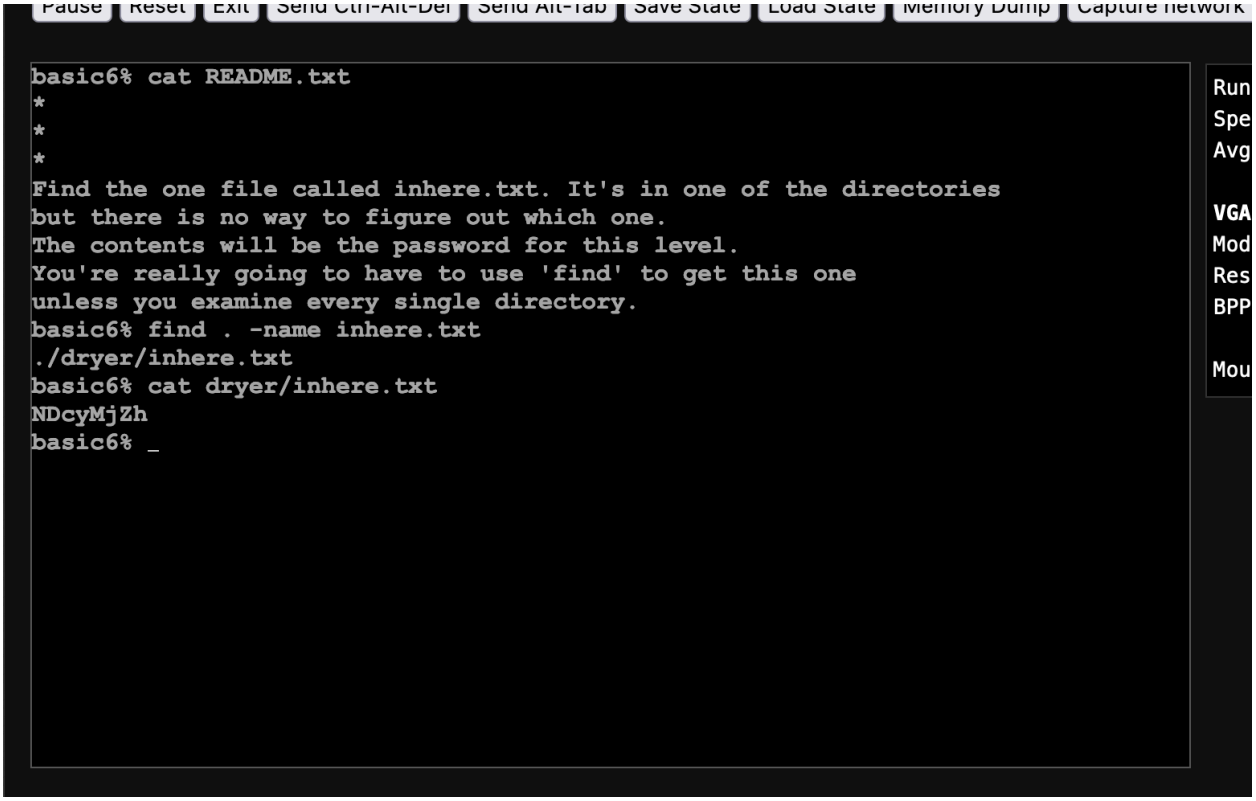


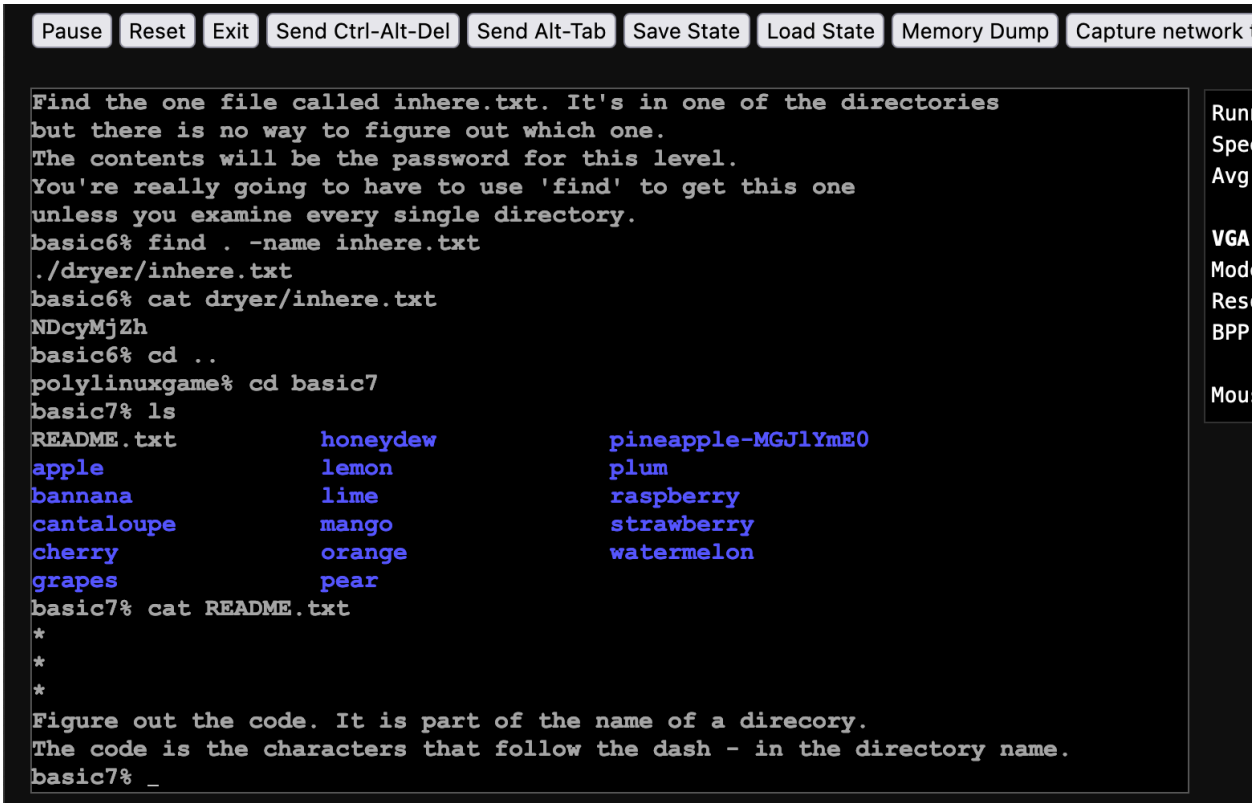
```
home% ls
polylinuxgame
home% cd polylinuxgame/
polylinuxgame% ls
basic1   basic2   basic4   basic6   basic8
basic10  basic3   basic5   basic7   basic9
polylinuxgame% cd basic1
basic1% ls
README.txt  inhere.txt
basic1% cat inhere.txt
YjIyZWYx
basic1% cd ..
polylinuxgame% cd basic2
basic2% ls
README.txt
basic2% cat README.txt
*
*
*
Display contents of the .inhere.txt file in this directory.
The contents will be the password for this level.
Note: This is a 'hidden' file.
basic2% cat .inhere.txt
MDcxM2Rm
basic2% _
```

```
polylinuxgame% cd basic2
basic2% ls
README.txt
basic2% cat README.txt
*
*
*
Display contents of the .inhere.txt file in this directory.
The contents will be the password for this level.
Note: This is a 'hidden' file.
basic2% cat .inhere.txt
MDcxM2Rm
basic2% cd ..
polylinuxgame% cd basic3
basic3% ls
README.txt     lufthansa.txt
basic3% cat README.txt
*
*
*
Display contents of the only .txt file in this directory except for README.txt.
The contents will be the password for this level.
basic3% cat lufthansa.txt
MTNkNWIz
basic3% _
```

```
*
*
*
Display contents of the only .txt file in this directory except for README.txt.
The contents will be the password for this level.
basic3% cat lufthansa.txt
MTNkNWIz
basic3% cd ..
polylinuxgame% cd basic4
basic4% ls
7up.txt          coke.txt         melloyellow.txt  sierramist.txt
README.txt       creamsoda.txt    mountaindew.txt  sprite.txt
birchbeer.txt    crush.txt        pepsi.txt
brie.txt         drpepper.txt     schweppes.txt
canadadry.txt    fanta.txt        seagrams.txt
basic4% cat README.txt
*
*
*
Display contents of the .txt file in this directory that is different than the o
thers.
The contents will be the password for this level.
basic4% cat brie.txt
NTA2M2Fh
basic4% _
```

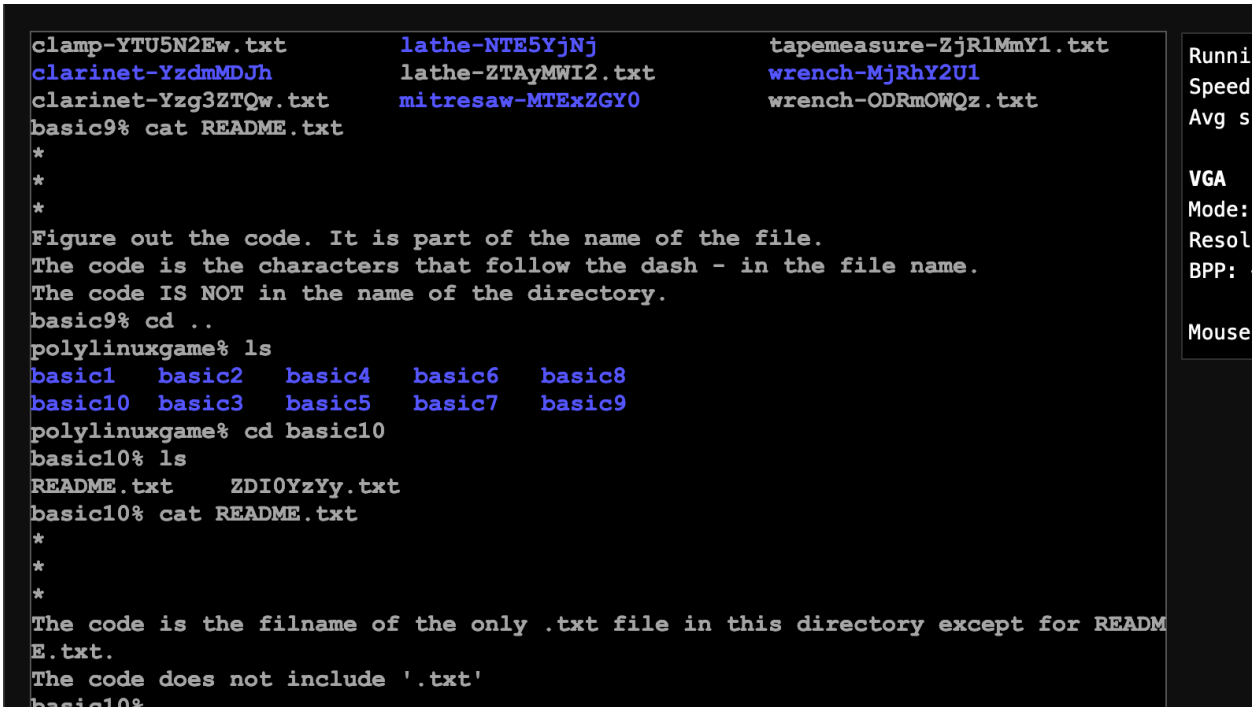
```
basic5% cat README.txt
*
*
*
Find the one file called inhere.txt. It's in the one directory
that is named differnetly than the others.
The contents will be the password for this level.
basic5% ls
README.txt    corgi         greyhound     poodle        terrier
beagle        dachshund     malamute      rottweiler
boston        dobermann     mastiff       saintbernard
chihuahua     greatdane     pitbull       spaniel
basic5% find . -name inhere.txt
./boston/inhere.txt
basic5% cd boston/
boston% ls
inhere.txt
boston% cat inhere.txt
NjE0OTEy
boston% _
```


Pause
Reset
Exit
Send Ctrl-Alt-Del
Send Alt-Tab
Save State
Load State
Memory Dump
Capture network
basic6% cat README.txt
*
*
*
Find the one file called inhere.txt. It's in one of the directories
but there is no way to figure out which one.
The contents will be the password for this level.
You're really going to have to use 'find' to get this one
unless you examine every single directory.
basic6% find . -name inhere.txt
./dryer/inhere.txt
basic6% cat dryer/inhere.txt
NDcyMjZh
basic6% _


Pause
Reset
Exit
Send Ctrl-Alt-Del
Send Alt-Tab
Save State
Load State
Memory Dump
Capture network
Find the one file called inhere.txt. It's in one of the directories
but there is no way to figure out which one.
The contents will be the password for this level.
You're really going to have to use 'find' to get this one
unless you examine every single directory.
basic6% find . -name inhere.txt
./dryer/inhere.txt
basic6% cat dryer/inhere.txt
NDcyMjZh
basic6% cd ..
polylinuxgame% cd basic7
basic7% ls
README.txt
apple
bannana
cantaloupe
cherry
grapes
honeydew
lemon
lime
mango
orange
pear
pineapple-MGJlYmE0
plum
raspberry
strawberry
watermelon
basic7% cat README.txt
*
*
*
Figure out the code. It is part of the name of a direcory.
The code is the characters that follow the dash - in the directory name.
basic7% _

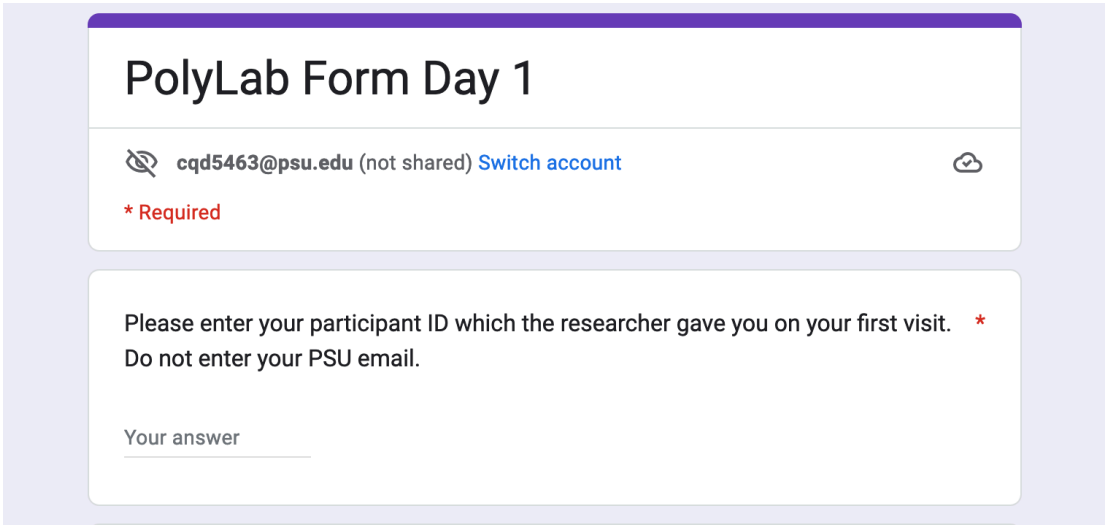
```
*
*
*
Figure out the code. It is part of the name of a direcory.
The code is the characters that follow the dash - in the directory name.
basic7% cd ..
polylinuxgame% cd basic8
basic8% ls
README.txt            guitar-NGMzOWU5       trombone-M2V1M2Y5
banjo-Y2JiM2Zk        mellophone-ODJiY2Yw   trumpet-NDRmYmQ0
bassoon-NjA3YjM2      piano-ZDExM2Yx        ukulele-OGRiYzM3
cello-NzExZDZk        poodle-OTc4YzQ3       viola-YzMxYWM4
clarinet-MWIyNGRh     saxophone-ZTU3ZWEw    violin-YmF1ZDll
flute-MjY0MmI5        souzaphone-YTJlNDFi
basic8% cat README.txt
*
*
*
Figure out the code. It is part of the name of a direcory.
The code is the characters that follow the dash - in the directory name.
The code is in the directory name that is different than the other directory nam
es.
```

```
The code is the characters that follow the dash - in the directory name.
The code is in the directory name that is different than the other directory nam
es.
basic8% cd ..
polylinuxgame% cd basic9
basic9% ls
README.txt                drill-OTYwNzJi            mitresaw-YzI3NGYw.txt
awl-MmEyZTF1              drill-ZjNiYTJh.txt        pliers-MWUyZDJi
awl-N2JkMmU5.txt          hammer-NWY5OTAx.txt       pliers-NTc3ZjY2.txt
beltsander-OWMxYmIx.txt   hammer-ZmM2MGI3           ratchet-MDhmYzVh.txt
beltsander-Yzg5Mjgz       jigsaw-ODBiNDdm.txt       ratchet-YjY1ZmEx
chisel-MzIzZTY4.txt       jigsaw-Zjk0YWVk           screwdriver-MGE2MDIw.txt
chisel-OWM1OWM2           ladder-NGZiZjBk.txt       screwdriver-MzU5MjVm
clamp-NTVjMDBk            ladder-ZjkxNTJk           tapemeasure-MWFjZjk4
clamp-YTU5N2Ew.txt        lathe-NTE5YjNj            tapemeasure-ZjRlMmY1.txt
clarinet-YzdmMDJh         lathe-ZTAyMWI2.txt        wrench-MjRhY2U1
clarinet-Yzg3ZTQw.txt     mitresaw-MTExZGY0         wrench-ODRmOWQz.txt
basic9% cat README.txt
*
*
*
Figure out the code. It is part of the name of the file.
The code is the characters that follow the dash - in the file name.
The code IS NOT in the name of the directory.
basic9% _
```



Day 1 Post Lab Questionnaire

This questionnaire was given to participants after they had completed the first lab session.



How mentally demanding was the task in comparison to other computer related tasks? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How physically demanding was the task in comparison to other computer related tasks? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How hurried or rushed was the pace of the task in comparison to other computer related tasks? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How successful were you in accomplishing what you were asked to do? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How hard did you have to work to accomplish your level of performance? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How insecure, discouraged, irritated, stressed, and annoyed were you? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How UNSTABLE was the situation that was presented in the scenario? *

	1	2	3	4	5	6	7	8	9	10	
Very Stable	○	○	○	○	○	○	○	○	○	○	Very Unstable

How COMPLEX was the situation that was presented in the scenario? *

	1	2	3	4	5	6	7	8	9	10	
Not Complex	○	○	○	○	○	○	○	○	○	○	Very Complex

How VARIABLE was the situation that was presented in the scenario? *

	1	2	3	4	5	6	7	8	9	10	
Not Variable	○	○	○	○	○	○	○	○	○	○	Very Variable

How STIMULATING was the situation that was presented in the scenario? *

	1	2	3	4	5	6	7	8	9	10	
Not Stimulating	○	○	○	○	○	○	○	○	○	○	Very Stimulating

To what degree did the situation in the scenario require you to CONCENTRATE YOUR ATTENTION? *

	1	2	3	4	5	6	7	8	9	10	
Low concentration needed	○	○	○	○	○	○	○	○	○	○	High concentration needed

To what degree did the situation in the scenario require you to DIVIDE YOUR ATTENTION? *

	1	2	3	4	5	6	7	8	9	10	
Low attention division	○	○	○	○	○	○	○	○	○	○	High attention division

While addressing the situation in the scenario, how much SPARE MENTAL CAPACITY would you say you had? *

	1	2	3	4	5	6	7	8	9	10	
Low spare mental capacity	○	○	○	○	○	○	○	○	○	○	High spare mental capacity

Please rate the QUANTITY OF INFORMATION that was presented in the scenario. *

	1	2	3	4	5	6	7	8	9	10	
Low quantity	○	○	○	○	○	○	○	○	○	○	High quantity

How FAMILIAR does the situation in the scenario feel to you? *

	1	2	3	4	5	6	7	8	9	10	
Not familiar	○	○	○	○	○	○	○	○	○	○	Very familiar

Submit Page 1 of 1 Clear form

Never submit passwords through Google Forms.

This form was created inside of psu.edu. Report Abuse

Google Forms

Day 2 Post Lab Questionnaire

This questionnaire was given to participants after they had completed the first lab session.

PolyLab Form Day 2

cqd5463@psu.edu (not shared) Switch account

* Required

Please enter your participant ID which the researcher gave you on your first visit. * Do not enter your PSU email.

Your answer

How mentally demanding was the task in comparison to other computer related tasks? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How physically demanding was the task in comparison to other computer related tasks? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How hurried or rushed was the pace of the task in comparison to other computer related tasks? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How successful were you in accomplishing what you were asked to do? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How hard did you have to work to accomplish your level of performance? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How insecure, discouraged, irritated, stressed, and annoyed were you? *

	1	2	3	4	5	6	7	8	9	10	
very low	○	○	○	○	○	○	○	○	○	○	very high

How UNSTABLE was the situation that was presented in the scenario? *

	1	2	3	4	5	6	7	8	9	10	
Very Stable	○	○	○	○	○	○	○	○	○	○	Very Unstable

How COMPLEX was the situation that was presented in the scenario? *

	1	2	3	4	5	6	7	8	9	10	
Not Complex	○	○	○	○	○	○	○	○	○	○	Very Complex

How VARIABLE was the situation that was presented in the scenario? *

	1	2	3	4	5	6	7	8	9	10	
Not Variable	○	○	○	○	○	○	○	○	○	○	Very Variable

How STIMULATING was the situation that was presented in the scenario? *

	1	2	3	4	5	6	7	8	9	10	
Not Stimulating	○	○	○	○	○	○	○	○	○	○	Very Stimulating

To what degree did the situation in the scenario require you to CONCENTRATE YOUR ATTENTION? *

	1	2	3	4	5	6	7	8	9	10	
Low concentration needed	○	○	○	○	○	○	○	○	○	○	High concentration needed

To what degree did the situation in the scenario require you to DIVIDE YOUR ATTENTION? *

	1	2	3	4	5	6	7	8	9	10	
Low attention division	○	○	○	○	○	○	○	○	○	○	High attention division

While addressing the situation in the scenario, how much SPARE MENTAL CAPACITY would you say you had? *

	1	2	3	4	5	6	7	8	9	10	
Low spare mental capacity	○	○	○	○	○	○	○	○	○	○	High spare mental capacity

Please rate the QUANTITY OF INFORMATION that was presented in the scenario. *

	1	2	3	4	5	6	7	8	9	10	
Low quantity	○	○	○	○	○	○	○	○	○	○	High quantity

How FAMILIAR does the situation in the scenario feel to you? *

	1	2	3	4	5	6	7	8	9	10	
Not familiar	○	○	○	○	○	○	○	○	○	○	Very familiar

Submit Page 1 of 1 Clear form

Never submit passwords through Google Forms

BIBLIOGRAPHY

A brief history of computer viruses. (n.d.). A Brief History of Computer Viruses. Retrieved April 3, 2023, from https://www.avg.com/en/signal/history-of-viruses

Arpanet. (2023, February 3). britannica. https://www.britannica.com/topic/ARPANET

Buildroot—Making embedded linux easy. (n.d.). Retrieved April 3, 2023, from https://buildroot.org/

copy. (2021). *V86* (d2f85b4). https://github.com/copy/v86

Cybersecurity Workforce Demand. (n.d.). National Institute of Standards and Technology. https://www.nist.gov/system/files/documents/2022/07/06/NICE%20FactSheet_Workforce%20Demand_Final_20211202.pdf

Harmon, T. (2016, September 14). Cyber security capture the flag (Ctf): What is it? *Cisco Blogs*. https://blogs.cisco.com/perspectives/cyber-security-capture-the-flag-ctf-what-is-it

Humphrey, David. (2021). *Browser-vm* (v1.0.2). https://github.com/humphd/browser-vm

Learning | linux journey. (n.d.). Retrieved April 3, 2023, from https://linuxjourney.com/lesson/the-shell

Morris worm. (n.d.). [Page]. Federal Bureau of Investigation. Retrieved April 3, 2023, from https://www.fbi.gov/history/famous-cases/morris-worm

Nicklaus A. Giacobe. (2013). *MEASURING THE EFFECTIVENESS OF VISUAL ANALYTICS AND DATA FUSION TECHNIQUES ON SITUATION AWARENESS IN CYBER-SECURITY* [The Pennsylvania State University]. https://etda.libraries.psu.edu/files/final_submissions/8031

Overthewire: Bandit. (n.d.). Retrieved April 3, 2023, from https://overthewire.org/wargames/bandit/

Practice labs | hands-on learning for digital & it skills. (n.d.). Retrieved April 3, 2023, from https://skills.practice-labs.com

Practice Labs is rated "Poor" with 1.9 / 5 on Trustpilot. (2023, March 17). Trustpilot. https://www.trustpilot.com/review/practice-labs.com

Ryan J. Kohler. (2018). *POLYMORPHIC HOMEWORK AND LABORATORY SYSTEM: AN INSTRUCTIONAL TOOL FOR COMBATING CHEATING* [The Pennsylvania State University]. https://etda.libraries.psu.edu/files/final_submissions/16872

Seed project. (n.d.). Retrieved April 3, 2023, from https://seedsecuritylabs.org/

Steve J. Cromity III. (2021). *POLYBANDIT: POLYMORPHIC LINUX COMMAND LINE SIMULATION* [The Pennsylvania State University]. https://etda.libraries.psu.edu/files/final_submissions/23730

What is kali linux? | kali linux documentation. (n.d.). Kali Linux. Retrieved April 3, 2023, from https://www.kali.org/docs/introduction/what-is-kali-linux/

What is Linux? (n.d.). Retrieved April 3, 2023, from https://www.redhat.com/en/topics/linux/what-is-linux

What is the CIA Triad and Why is it important? (n.d.). Fortinet. Retrieved April 3, 2023, from https://www.fortinet.com/resources/cyberglossary/cia-triad

ACADEMIC VITA

EDUCATION:

The Pennsylvania State University University Park, PA
- BS Cybersecurity Analysis and Operations May 2023
- National Security Agency Certificate
- Schreyer Honors College

PROJECT EXPERIENCE:

Schreyer Honors College Research and Thesis
- Created a lab environment for students to learn basic Linux commands
- Conducted social science research on the effectiveness of new teaching methods

RSM Campus Automation Challenge 2021
- Won most creative automation bot
- As a team, we developed a bot that was capable of collecting and displaying information on a given domain and subdomains.

Hack PSU Vanguard Challenge
- Our project won 2nd place in the Vanguard challenge at Hack PSU 2020.
- I worked in a team that developed a GroupMe chat bot in Python to help college students do their taxes.

National Cyber League Competition, *Team Leader*
- As a team we worked together to solve a verity of capture the flag objectives with the use of Wireshark, Nmap, log analysis, etc.

EXPERIENCE:

Penn State, *Learning Assistant (LA)* Jan 2020-Present
- Classes: IST 140, IST 210, IST 220, IST 242, Cyber 366, IST 454
- Hold office hours for students
- Grade assignments
- Answer student questions and assist with teaching: Java, SQL, malware analytics, cyber forensics, and networking

Social Impact Capital (SIC), *Intern* Mar 2020-April 2020
- Performed technical due diligence on ML investment opportunities for SIC
- Evaluate and test software applications and SDKs
- Internship ended early due to the pandemic

GDS, *Lead High School Technology Intern* June 2018-Aug 2019
- Managed the workload of five other interns
- Developed Unix shell scripts for managing computers
- Implemented device management systems on Apple devices (Jamf)
- Troubleshot networking and server equipment
- Worked on repairing laptops and desktop computers.

SKILLS:

Languages:	Java, Python, SQL, HTML, CSS, JavaScript, Bash, Arduino
Operating Systems:	Linux, Kali Linux, macOS, Windows 11/10/8/7/Server
Applications:	ArcGIS, PyCharm, IntelliJ, NetBeans, MS SQL Server, VS Code, Wireshark, Nmap, Git, Zsh, Emacs, Jamf, Active Directory, Microsoft Visio, Metasploit, Splunk

ACTIVITIES:

Member Gamma Tau Phi, IST Honor Society Aug 2020-Present