Highlighting Vulnerabilities in a Genomics Biocybersecurity Lab Through Threat Modeling and Security Testing

Jared Sheldon[a], Isabelle Brown-Cantrell[b], Patrick Pape[c] and Thomas Morris[d]
Center for Cybersecurity Research and Education, University of Alabama in Huntsville, Huntsville, Alabama, U.S.A.

Keywords: Biocybersecurity, Cybersecurity, Genomics, Threat Modeling, ATT&CK, TTPs, STRIDE.

Abstract: Biocybersecurity, a specialty field applying modern cybersecurity developments to the bioeconomy, is garnering progressively more attention as concerns increase over the protection of bioeconomic data generated each year. Genomic data is a key data type that falls under the bioeconomy umbrella and can be protected health information, intellectual property, or research data, depending on the use case. To increase understanding of cybersecurity for genomic lab environments, a biocybersecurity laboratory was set up and threat modeling was conducted on it using the STRIDE threat modeling methodology. Potential attack techniques were then mapped using the MITRE ATT&CK enterprise matrix and attack trees were generated to sequentially show the steps of these attacks. Going a step further, the initial steps of an attack tree were attempted against a DNA sequencer in the biocybersecurity lab. While the results of this testing did not yield an exploitable vulnerability that could be used to further test the attack tree techniques, lessons learned along the way can be taken into account by future research projects pursuing similar goals.

1 INTRODUCTION

Genomic data is highly important and the environments that generate this data have unique characteristics that must be accounted for when seeking to protect it. Whether the genomic data is Protected Health Information (PHI), Intellectual Property (IP), or research project data, its loss would present a significant loss of time, money, and potentially privacy for individuals if the genomic data is from a human.

Each genome sequenced is the result of laboratory technicians spending time moving a DNA sample through a series of laboratory machines that each prepare the DNA sample in a different way before arriving at the DNA sequencer. After these preparatory steps have been completed and quality of the sample has been assured, the laboratory technicians use the DNA sequencer to generate digital data from the physical sample. Throughout this process, consumables and time have been used to prepare the sample and generate the resulting digital data. All of this investment can be made moot if the resulting data is lost or corrupted. This loss of economic investment is only deepened if the genomic data lost is IP, as would be the case for genetically modified crops or biopharmaceuticals, since such products also represent an investment in research and development time.

Aside from economic investment being lost, individual privacy can be impacted if the genomic data lost was a person's PHI. The impact that exposure of this kind of PHI in a data breach scenario can have is only worsened by the fact that the relevant individual's family is also affected. This trait of genomic data introduces unique privacy concerns, on top of the concerns already at play, since genomic data is PHI that never or rarely changes for affected parties. Unlike a credit card number exposed in a data breach, a person whose genomic data has been exposed cannot simply change their data. For cybersecurity and privacy, this coupled with the data's ability to affect entire family trees means that genomic data should be protected as PHI for as long as it is stored.

This next section of this paper covers details regarding a biocybersecurity lab (BCL) created to facilitate biocybersecurity research. The following section then discusses the STRIDE threat modeling effort conducted on the BCL and attack mappings generated with the gathered threat modeling insights. Network scans demonstrating the initial reconnaissance phase

[a] https://orcid.org/0009-0009-7909-4217
[b] https://orcid.org/0009-0004-8820-6448
[c] https://orcid.org/0009-0005-4922-4026
[d] https://orcid.org/0000-0002-4854-5419

Sheldon, J., Brown-Cantrell, I., Pape, P., Morris and T.
Highlighting Vulnerabilities in a Genomics Biocybersecurity Lab Through Threat Modeling and Security Testing.
DOI: 10.5220/0013523800003979
In *Proceedings of the 22nd International Conference on Security and Cryptography (SECRYPT 2025)*, pages 626-631
ISBN: 978-989-758-760-3; ISSN: 2184-7711
Copyright © 2025 by Paper published under CC license (CC BY-NC-ND 4.0)

of a potential attack targeting an Illumina NovaSeq 6000 are then reviewed. Results and considerations are then covered, followed by future work ideas and conclusions that were drawn from this research endeavor.

2 BIOCYBERSECURITY LABORATORY

For the past four years, we have partnered with the HudsonAlpha Institute for Biotechnology, a local genomic sequencing laboratory campus, to conduct research into the genomic threat landscape. Over the past year, the sequencing laboratory has created a hands-on, modular biocybersecurity laboratory (BCL) to spearhead crucial research into the area. Through our partnership, we were given access to the BCL to conduct the threat modeling exercises and network scans discussed in the succeeding sections.

2.1 Laboratory Setup

The BCL currently consists of a 1,224-square foot lab space containing devices comprising the first two stages of the genomic data life cycle: creation and storage. This includes a Laboratory Information Management System (LIMS) to document sample intake, prescribe a psuedoidentifier, and keep track of the sample as it is sequenced in the lab. Next, the BCL contains genomic devices that handle DNA extraction, DNA fragmentation, library preparation, and quality control before the sample is fed to a genomic sequencer. To provide a well-rounded laboratory environment for testing, the BCL has devices from multiple major genomic device manufacturers such as PacBio, Illumina, Tecan, and Agilent Technologies. These fully installed and operational devices in the BCL can be seen in Figure 1.

Figure 1: BCL genomic sequencing devices installed and operational.

2.2 Laboratory Purpose

The BCL was created to provide a unique opportunity for students, researchers, and organizations to carry out both technical research projects and educational learning experiences. Several camps and trainings have been conducted in the BCL aimed at teaching high school and undergraduate college students about genomic sequencing devices, their purpose in the life cycle, how to secure them, and what happens to the sequence data created by the devices. Additionally, the BCL is set up to allow organizations to test the application of cybersecurity and privacy standards and frameworks currently being developed.

3 THREAT MODELING

To better understand the organization and capabilities of the BCL, a thorough cybersecurity threat model was created and iterated over. The first step in threat modeling the BCL environment was to discuss the laboratory configuration and data flows with BCL staff and create a series of diagrams documenting the lab. Next, a STRIDE analysis was conducted against the components and data flows documented in the diagrams. These steps and their outcomes are detailed below.

3.1 Diagramming

The diagramming process began with a series of in-person tours at the local genomic sequencing laboratory campus. These tours included a detailed walkthrough of how each device they own fits into the genomic data life cycle, how lab technicians interact with each device, and the overall process of going from obtaining a physical sample to having detailed analytical results of the genome occurs. Physical and network segmentations were discussed to determine the presence of inherent security boundaries for the diagrams. After the in-person tours, weekly meetings were held with BCL staff members to discuss the lab setup and devices further and to address any questions that came up as the diagrams were developed.

After determining the basic components, flows, and security boundaries that needed to be present in the diagram, it was essential to determine a common notation to use to ensure ease of readability. The notation developed and documented in MITRE's Playbook for Threat Modeling Medical Devices was utilized (Bochniewicz et al.,). This notation has six unique components: processes, trust boundaries, external entities, data stores, users, and data flows.

To ensure reader understanding, additional images and icons were used in the diagrams, such as a genomic sequencer icon within the component box for the genomic sequencer. Color was also utilized to show a separation between the elements of the BCL and their underlying trust boundaries. A detailed breakdown of the wet laboratory within the BCL can be seen in Figure 2.

3.2 STRIDE Analysis

Given the complexity of the DFDs, it became important to prioritize the data flows for which threats would be modeled. Doing this would allow for the threat modeling effort to focus on all of the components and the highest value data flows to keep the effort more focused. To accomplish this prioritization, we identified the data flows that were of high value either due to the value of the data sent over the data flow or due to the general criticality of the data flow to laboratory operations. To ensure accuracy, these high value data flows were presented to specialists from our sequencing lab partner to confirm that the chosen data flows were where time spent threat modeling would provide the most benefit to a genomics lab.

Once this list of high value data flows was confirmed, a STRIDE analysis was conducted. This analysis used the STRIDE threat modeling methodology to elicit the spoofing, tampering, repudiation, information disclosure, denial of service, and elevation of privilege threats applicable to all components within the threat model as well as those applicable to the high value data flows (Shostack, 2014). To maintain consistency throughout the process of identifying threats, Table 3 from the Playbook for Threat Modeling Medical Devices (Bochniewicz et al.,) was used to provide a basis for which STRIDE elements were applicable to which types of components and data flows. This increased the STRIDE analysis speed and resulted in the identification of over two hundred threats across the genomic lab threat model.

3.3 Attack Mapping

After enumerating the possible threats and mitigations for each lab component, it was essential to map the identified threats to a well-known, standard framework. For this purpose, the MITRE Adversarial Tactics, Techniques, and Common Knowledge (MITRE ATT&CK) framework was chosen (MITRE, n.d.). The MITRE ATT&CK framework consists of 14 tactic categories with over 200 individual techniques. These techniques range from open-source intelligence gathering to utilizing a command and control channel to exfiltrate data. The abundance of potential techniques that are highly specific allows for detailed mappings between STRIDE threats and ATT&CK techniques to be possible. The result of this mapping can be seen in Table 1.

To create the mappings seen in Table 1, the threat descriptions created during the STRIDE analysis were utilized. The team evaluated the descriptions altogether to determine the tactic category and individual technique for the mappings, as well as the individual components of the description, effectively creating an attack chain of ATT&CK techniques.

4 NETWORK SCANS

Building off of the attack mapping performed, the biocybersecurity lab was leveraged as a target environment for network scans and tests. The device selected for these scans and tests was an Illumina NovaSeq 6000 no longer used in production environments. This device was deployed in the biocybersecurity lab by the sequencing lab partner, the device owner, and access to a virtual machine on the BCL network was used to conduct the following tests.

The next scans performed were TCP and UDP Nmap (Lyon, n.d.) scans of the sequencer with the goal of determining open port numbers. Once the open port numbers were identified, a series of scans were performed to find what Nmap guessed as the operating system and to have Nmap identify the services running on those open ports. The information from these scans informed the types of scans and tests performed next. The SYN scan results can be seen in Figure 3.

The most interesting service identified from these scans was an HTTP server. This HTTP server was heavily targeted in a series of numerous tests. These tests included attempts to leverage HTTP verbs using cURL (curl,) to determine if any would yield interesting results. The next tests also used cURL and were attempts at directory traversal attacks through manipulating the URL targeted. Another round of tests included banner information gathering through a variety of tools in an attempt to determine more information about the running HTTP server. No interesting results were found in these tests.

Nikto, a web application vulnerability scanner (Sullo and Lodge, n.d.), was used to scan the web server, but still no useful information was returned. Gobuster (OJ, n.d.) was used to try enumerating the directories on the HTTP server using the SecLists combined_directories.txt wordlist (Miessler et al., n.d.). No results were returned from this enu-

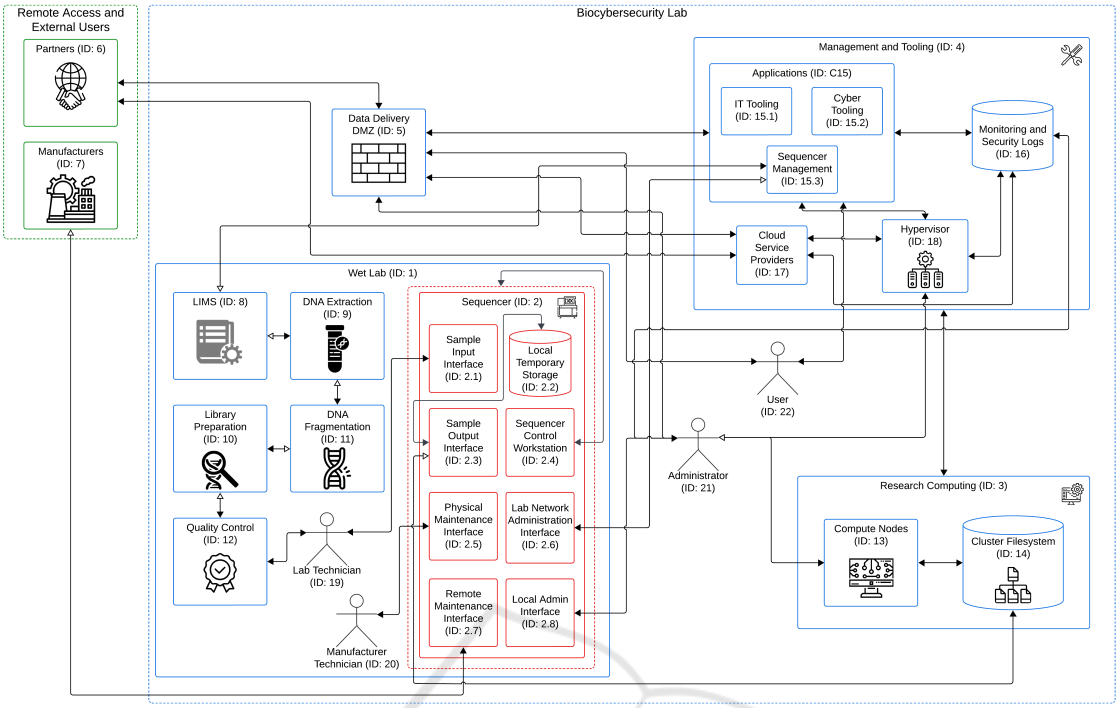

Figure 2: Detailed BCL DFD.

```
Not shown: 65523 closed tcp ports (reset)
PORT      STATE SERVICE       VERSION
135/tcp   open  msrpc         Microsoft Windows RPC
139/tcp   open  netbios-ssn   Microsoft Windows netbios-ssn
445/tcp   open  microsoft-ds?
5357/tcp  open  http          Microsoft HTTPAPI httpd 2.0 (SSDP/UPnP)
|_http-title: Service Unavailable
|_http-server-header: Microsoft-HTTPAPI/2.0
29644/tcp open  http          Microsoft HTTPAPI httpd 2.0 (SSDP/UPnP)
|_http-title: Bad Request
|_http-server-header: Microsoft-HTTPAPI/2.0
49664/tcp open  msrpc         Microsoft Windows RPC
49665/tcp open  msrpc         Microsoft Windows RPC
49666/tcp open  msrpc         Microsoft Windows RPC
49667/tcp open  msrpc         Microsoft Windows RPC
49668/tcp open  msrpc         Microsoft Windows RPC
49669/tcp open  msrpc         Microsoft Windows RPC
49680/tcp open  msrpc         Microsoft Windows RPC
MAC Address: 00:90:FB:5E:42:49 (Portwell)
Device type: general purpose
Running: Microsoft Windows 10
OS CPE: cpe:/o:microsoft:windows_10
OS details: Microsoft Windows 10 1507 - 1607
Network Distance: 1 hop
Service Info: OS: Windows; CPE: cpe:/o:microsoft:windows

Host script results:
| smb2-time:
|   date: 2024-10-11T15:47:04
|_  start_date: 2024-09-10T18:02:37
|_nbstat: NetBIOS name: NOVASEQ, NetBIOS user: <unknown>, NetBIOS MAC: 00:90:fb:5e:42:49 (Portwell)
| smb2-security-mode:
|   3.1.1:
|_    Message signing enabled but not required
|_clock-skew: -2s

OS and Service detection performed. Please report any incorrect results at https://nmap.org/submit/ .
```

Figure 3: SYN Scan of DNA Sequencer.

Table 1: STRIDE-to-TTP Mapping Results.

Component Name	ID	S	T	R	I	D	E
Wet Lab	1	T1078	T1565	T1070	T1590	T1499	
Sequencer	2		T1091		T1040	T1499	T1068
Sample Input Interface	2.1		T1059	T1556		T1529	T1068
Local Temporary Datastore	2.2		T1565		T1005	T1529	
Sample Output Interface	2.3				T1041		
Sequencer Control Workstation	2.4		T1565		T1005	T1529	T1068
Physical Maintenance Interface	2.5		T1542	T1485		T1529	T1547
Lab Network Administration Interface	2.6		T1071	T1485	T1003	T1529	T1569
Remote Maintenance Interface	2.7		T1565	T1485		T1529	T1563
Research Computing Environment	3	T1078	T1222	T1485	T1005	T1499	T1548
Management and Tooling	4	T1078	T1070	T1485	T1005	T1529	T1078
Data Delivery DMZ	5	T1199	T1222	T1485	T1005	T1498	
Partners	6	T1199		T1021	T1537		
Manufacturers	7	T1199		T1070			
LIMS	8		T1600	T1485	T1005	T1499	T1134
DNA Extraction	9		T1565		T1005	T1499	
DNA Fragmentation	10		T1565		T1005	T1499	
Library Preparation	11		T1059		T1040	T1499	
Quality Control	12		T1600		T1040	T1499	
Compute Nodes	13	T1078	T1195	T1485	T1005	T1499	T1053
Cluster Filesystem	14		T1222		T1005	T1499	
APplications and Services	15	T1078	T1195	T1485	T1005	T1489	T1611
IT Tooling	15.1	T1078	T1195	T1485	T1005	T1489	T1611
Cyber Tooling	15.2	T1078	T1195	T1485	T1005	T1489	T1611
Sequencer Management	15.3	T1078	T1195	T1485	T1005	T1489	T1611
Monitoring and Security Logs	16		T1070		T1005	T1489	
Cloud Service Providers	17	T1078		T1485		T1489	
Hypervisor	18		T1564		T1489		
Lab Technician	19	T1078		T1485	T1650		
Manufacturer Technician	20	T1078	T1542	T1485	T1056		
Administrator	21	T1078		T1485	T1650		
User	22	T1078		T1485	T1650		

meration attempt. Additional cURL tests were performed with the goal of getting more informative responses from the HTTP server, such as user agent spoofing and specifying the allow unsafe option, but responses to these requests were no more elucidating. Network fuzzing using (xmendez, n.d.) was then conducted using the SecLists combined wordlist for HTTP server testing to find network requests that could be sent to the HTTP server to get more interesting information in responses. Analysis was performed on the number of characters in the responses returned by the server during this testing and found nothing of note.

4.1 Results and Considerations

Ultimately, the scans and tests targeting the DNA sequencer did not find an exploitable vulnerability. However, there are still lessons to be learned from the effort regarding technical information acquisition and self-reliant device deployment. These were issues that presented themselves throughout the research project and were difficult to overcome.

During the research project, it was difficult to locate full technical details about the target sequencing device from the manufacturer. The most useful technical information is held behind a pay wall and is not readily available for the public, making research efforts such as this more difficult. Accessing this information was not easier for the local genomic sequencing laboratory that contributed to the project. Having access to device documentation which is more technical than what is publicly available would have been beneficial to this research project and future projects.

In reviewing the data collected from the tests and scans performed on the DNA sequencer, the research

team believes that some of the sequencer's behavior – such as the error responses from the HTTP servers – may be attributable to not being fully deployed in a production environment by a manufacturer technician. Technical documentation defining expected behavior is necessary to confirm our suspicions. Future projects would benefit from engagement from the device manufacturer to fully deploy target sequencer devices.

5 FUTURE WORK

Further research into the overall security posture of genomic-specific devices, such as DNA sequencers and other wet lab devices, would contribute greatly to the nascent, yet maturing, field of biocybersecurity. Vulnerability assessments of the vast number of device models used in wet labs, starting with those devices that represent the largest market share, would improve sequencing lab trust in the devices that they connect to their networks and rely on for the production of DNA sequence data. While unachievable by academia alone, the creation of Manufacturer Usage Descriptions (MUD) stands to benefit genomics labs as network access for specialty devices such as the DNA sequencer could be reduced to only those connections that the device strictly needs.

6 CONCLUSIONS

Working together with a local genomic sequencing lab has provided valuable insight into the inner workings of genomic labs. Through consistent communication with the lab, the threat model was able to be iteratively developed. As new characteristics of the network were discovered through tours or interviews, previous threat modeling steps were revisited and adapted as needed to fit the new information. This led to the model's fidelity increasing over the course of the project.

The most interesting finding from the threat model to note is that device manufacturers or vendors may require direct access to their deployed devices in their customers' networks. For example, a DNA sequencer may require that it be reachable from the manufacturer for the purposes of updates and maintenance. This presents a cybersecurity concern as network administrators must take this into account when designing firewall rules or monitoring network traffic. In the case of a DNA sequencer, it is also worth noting that some manufacturers or vendors may use remote maintenance software to access the PC workstation attached to the sequencer when performing maintenance. Manufacturers or vendors may also send a maintenance technician to the sequencing lab's campus in-person to perform maintenance such as updates, depending on the situation.

Access to the BCL provided access to devices for research that otherwise would have been too cost prohibitive to conduct due to device prices. This allowed us to conduct network scans and tests in an environment that can be expanded over time. Leveraging a non-production lab allowed us to conduct network scans and tests without concern for harming an ongoing sequencing workflow.

Although a vulnerability was not discovered from the tests and scans, the effort has shown that more detailed technical documentation from manufacturers would assist research efforts and that future efforts may benefit from manufacturer-guided deployments in biocybersecurity testing labs. These guided deployments would ensure that the device is properly configured and that its network services are fully operational.

ACKNOWLEDGEMENTS

We acknowledge the HudsonAlpha Institute for Biotechnology for collaborating with this research effort and providing insight into a real-world genomics lab. Their involvement allowed for this research effort to have access to real-world genomics equipment and a functional environment to conduct threat modeling on and to perform scans and tests on.

REFERENCES

Bochniewicz, E., Chase, M., Coley, S. C., Wallace, K., Weir, M., and Zuk, M. Playbook for threat modeling medical devices.

curl. Official curl website.

Lyon, G. F. Nmap official site.

Miessler, D., Haddix, J., Portal, I., and g0tmi1k. Seclists github repository.

MITRE. Mitre att&ck enterprise matrix.

OJ. Gobuster github repository.

Shostack, A. (2014). *Threat modeling: Designing for security*. Wiley.

Sullo, C. and Lodge, D. Nikto download page.

xmendez. Wfuzz github repository.