Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2024) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

*The 5th International Workshop on Data- Driven Security 2024*
April 23-25, 2024, Hasselt, Belgium

# Unsafe at any Bandwidth: Towards Understanding Risk Factors for Ransomware in Higher Education

Logan Suarez[a], Dakhilallah Alshubrumi[a], Tj O'Connor[a,b] and Sneha Sudhakaran[a,b]

[a] *Department of Electrical Engineering and Computer Science, Florida Institute of Technology, Melbourne, FL, USA*

[b] *Faculty, L3Harris Institute for Assured Information, Florida Institute of Technology, Melbourne, FL, USA*

**Abstract**

United States higher education institutions host an assortment of services that are accessible via public IP addresses. The wide variety of network services and the important personal and institutional data stored on such services make higher education institutions particularly desirable targets for attackers. This study analyses the vulnerabilities found through Shodan scans on these networks, in conjunction with institutional characteristics data taken from the National Center for Education Statistics (NCES), to examine correlations between an institution's characteristics and the vulnerabilities found on its networks. By exploring this data, the study aims to bring awareness to the current state of higher education institution network security and determine vulnerability trends between certain institutions. Our analysis reveals that most institutions have many medium impacts but highly exploitable vulnerabilities, with most being on Apache HTTP servers. We also present that the most significant indicators of an institution's vulnerability are its enrolment and yearly total expenses. We investigate how smaller institutions have lower numbers of vulnerabilities, but their vulnerabilities have the potential for higher impact. We conclude that there is a significant chance of ransomware risk in US higher educational institutions.



*Keywords:* Shodan, Vulnerability Assessment, Higher Education

## *1. Introduction*

In the last two decades, the complexity of higher education computer networks has grown exponentially to accommodate the vast array of information technology applications and services. Due to the diversity of users and access means, these networks are more publicly accessible than similarly sized commercial organizations [12]. These always-available, publicly accessible, complex networks present substantial security challenges. The COVID-19 pandemic and the shift to hybrid and online learning modalities further complicated this challenge. In response to the pandemic, network administrators raced to make their networks worldwide accessible to students, faculty, and

staff. However, this push to expand the network and ensure worldwide accessibility exposed academia to ransomware groups. A survey of 5,600 IT professionals from 410 higher education institutions across 31 countries identified a surge of higher education-targeted ransomware in 2021. Our work mainly targets understanding the risk of academic institutions to ransomware groups. Previous works have examined vulnerability exposure at specific institutions [8], actively scanned a set of academic networks [11], and constructed risk assessment models for vulnerabilities [10]. In contrast, our work seeks to correlate the differences in educational institutions to their vulnerability to ransomware attacks by producing and analyzing the largest dataset of vulnerabilities. We leverage the publicly available Shodan search engine to construct our dataset to ensure reproducibility so that further researchers can build upon work. For reproducibility, we open-source all our work at <*https://redacted*>. This paper presents our findings that identify and correlate the risk factors for ransomware for higher education institutions. We construct a dataset of 744 higher education institutions that own autonomous system numbers (ASNs) and catalog their ransomware exposure from the Shodan [1] search engine results. We then analyze these results in the context of their institution's background using the Integrated Post-secondary Education Data System (IPEDS)[3]. Our work uncovers a heterogeneous disposition of vulnerabilities on publicly accessible services. Further, our work examines the most relevant correlations between the characteristics of an institution and exposure to high-risk vulnerabilities.

**This paper makes the following contributions:**

- We assessed the risk of 744 higher education institutions' public-facing IP addresses using Shodan, uncovering 1,306,816 vulnerabilities across 150,255 services, with 19.17% categorized as high-risk and an average disclosure age of over three years.
- We analyse the correlations between a school's institutional characteristics and the quantity and risk category of vulnerabilities present on a school's public-facing IP addresses. Our work reveals two key findings: (1) institutions with more students and greater total expenses had a higher correlation with the number of vulnerabilities, and (2) institutions with fewer students and total expenses had vulnerabilities with higher impact.
- To allow others to build on our work, we publish all our scripts at *https://redacted* for others to repeat our experiments. We will be publishing the work publicly once this work is accepted.

Section 2 motivates our work and explores network scanning and vulnerability assessment in the context of ransomware attacks. Section 3 presents an overview, identifying the critical challenges and data sources for examining the problem. Section 4 provides information on how we addressed the challenges of project data collection, correlating data sources, and risk factors. Section 5 details our analysis and results of the collected data. Section 6 presents our detailed results. Section 7 concludes the study.

## 2. Related Work and Background

In 2021, eighty-eight education sector organizations suffered ransomware attacks, twenty-six from colleges and universities [11]. Forty-four of these attacks publicly leaked sensitive information about students, faculty, and staff. Academic institutions are an attractive target for ransomware groups due to the wealth of sensitive data about students, finances, operations, and research partnerships [6]. The lack of resources and skilled administrators also challenges universities to defend their complex networks adequately. Further, it proves valuable to examine the impact of the 2021 Ransomware attack on Lincoln College. Established in 1865, Lincoln College shut down in May 2022 after a ransomware attack. Despite paying a $100,000 ransom, Lincoln College could not access its data for three months [9]. This challenge complicated an already struggling university during the difficulties of the COVID-19 pandemic.

### 2.1 Vulnerability and Risk Data Sources

***National Vulnerability Database:*** We leveraged the National Vulnerability Database (NVD) to quantify and understand risk in our work [4]. This NIST-developed database catalogs cyber vulnerabilities with a unique Common Vulnerability Exposure (CVE) identifier. NIST quantifies each CVE ID based on metrics, including the vulnerabilities' severity, exploitability, and impact. The CVE ID and associated score serve as a reference for tracking and assessing the risk of vulnerabilities. In addition to the common vulnerability exposure, NIST also quantifies vulnerabilities based on the complexity and availability of a remedy. In quantifying the risk and exposure across university networks, our work prioritized the legacy CVSS 2 scoring version, as 30% of vulnerabilities were

older than the 2015 CVSS 3 scoring metric. In understanding the organizational characteristics of universities, we leveraged the Integrated Postsecondary Education Data System (IPEDS.) This database records National Center for Education Statistics (NCES) survey data [3]. These surveys document institutional characteristics, including the institution name, location, institution type, campus setting, tuition, fees, and the staff-to-student ratio. At the time of our study, the IPEDs database relied on surveys from 2019-2020, providing the organizational characteristics from 4,186 higher education institutions in the United States.

***American Registery for Internet Numbers:*** We also leveraged the American Registry of Internet Numbers (ARIN) Whois database to identify higher education institutions' networks and Internet protocol (IP) addresses [2]. The publicly accessible ARIN Whois database stores information correlating domain names and organizations to internet routing registries. Further, the ARIN Whois database provides a method to query specific organizations and their assigned Autonomous System Numbers (ASNs) to retrieve internet protocol (IP) address ranges. Finally, the Shodan vulnerability search engine routinely scans public-facing IP addresses to document the current security state of networked systems [1].

***Shodan Vulnerability Search Engine:*** Shodan indexes the vulnerabilities, misconfigurations, and unprotected systems by IP address. Shodan allows network administrators to understand their network exposure passively without performing active network scanning. Further, Shodan enables researchers to query the database with advanced filters, including searching by IP address, network, specific vulnerability, or CVE identifier [1]. While Shodan is freely available, it requires creating a login and agreeing to the terms of service. Shodan requires a paid plan, or an account created with an academic email address to perform continuous automated scanning and analysis.

## 3. Challenges

In developing a dataset to understand higher educational institutions' characteristics and risk of cyber vulnerabilities, we had to solve three critical challenges.

1. **Identify all public-facing systems for universities:** To properly understand the risk, vulnerability, and exposure across higher education institutions, we first needed to uncover the IP addresses of all publicly facing network resources at an institution.
2. **Catalog vulnerabilities and risk:** Next, we needed to catalog the currently exposed vulnerabilities and their risks across the higher education IP addresses.
3. **Analyze trends to correlate and uncover risk features:** Finally, we needed to analyze our data to identify correlations and understand risk causality across different higher educational institutions' characteristics. 4. **Proposed Methodology**

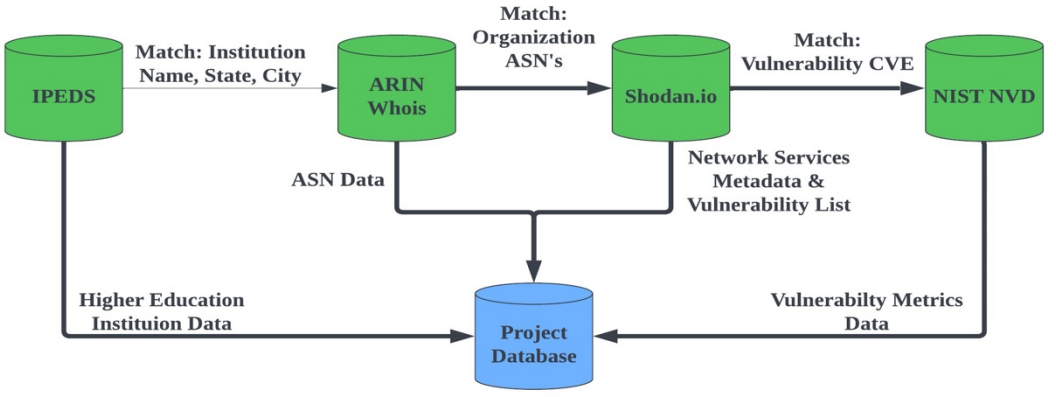



Fig. 1. Data collection and matching process

### 3.1 College Characteristic Data Collection

To analyse the data relevant to our problem, we collected and matched information from the ARIN, NIST, Shodan, and IPEDs databases and stored it in a PostgreSQL database. We created a record of each institution, cataloguing their public services and vulnerability exposure. For others to build upon our work, we provide the Python3 scripts for collecting, matching, and inserting this data into a PostgreSQL database for analysis. Figure 1 illustrates this

process. We leveraged the IPEDS dataset to acquire characteristics data about higher education institutions. This dataset was available for download as a Microsoft Access database dump, which we then converted into a PostgreSQL database. Further, we developed a set of specific queries for the IPEDS database describing the institutional characteristics we wanted to explore. We leveraged the IPEDS dataset to acquire characteristics data about higher education institutions. This dataset was available for download as a Microsoft Access database dump, which we then converted into a PostgreSQL database. Further, we developed a set of specific queries for the IPEDS database describing the institutional characteristics we wanted to explore.

### 3.2 College Public Ip Address

After retrieving the records of United States colleges from the IPEDS dataset, we leveraged the ARIN Whois [2] API to identify the ASNs related to each of these organizations. Further, this approach allowed us to identify and match blocks of public IP addresses to registered college organizations by querying for colleges via organization name.

**Table 1. Vulnerability metrics score percentages**

| Metric | Low | Medium | High |
|---|---|---|---|
| Severity | 5.6% | 74.7% | 19.7% |
| Exploitability | 4.8% | 5.2% | 90.0% |
| Impact | 60.4% | 36.3% | 3.3% |

One challenge with querying colleges using their institution name was that the naming conventions of many organizations were inconsistent, so we had to correct this by identifying institutions represented by multiple names (e.g., Georgia Institute of Technology vs. Georgia Tech). We ensured the integrity of our dataset by adding records only where the ASN's organization location from the ARIN dataset reached the institution's location in the IPEDS dataset.

### 3.3 Services and Vulnerability Scanning

Finally, by providing ASN filters, we queried institutional and their associated IP subnets for exposed services through Shodan's API. We downloaded and stored this data as JSON records. These JSON records contained a wide variety of data on the services running at each IP address, including device, product, and vulnerability information detected through metadata. Since Shodan runs regular scans, we identified that the service/vulnerability information was at most a week old. After collecting these results, we inserted the records into the project database. The Shodan results further allowed us to explore relevant vulnerability metrics information by correlating the exposure to the National Vulnerability Database using the CVE ID of all vulnerabilities detected in the Shodan scans.

## 4.Results

Utilizing the results from Shodan and the IPEDS dataset, we obtained statistics on the type and count of vulnerabilities found in the networks. We then studied how these statistics relate to the institution's characteristics.

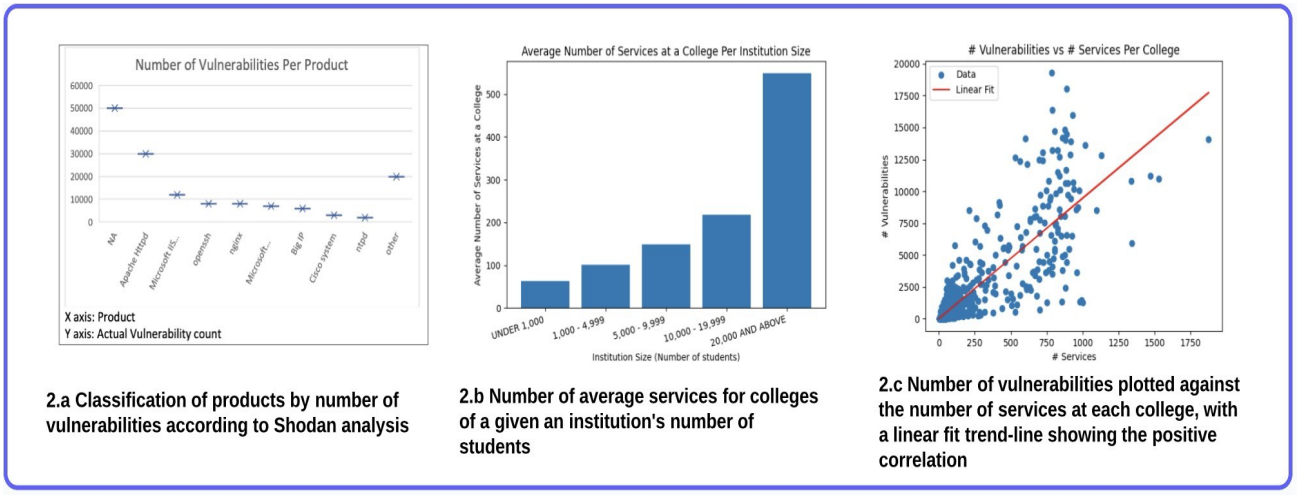
### 4.1. Total Vulnerabilities Uncovered

Among the 744 institutions studied, 803 ASNs hosted active services, with 27,534 of these services displaying vulnerabilities. These vulnerabilities amounted to 1,306,816, categorized as 5.6% low severity, 74.7% medium severity, and 19.7% high severity, as detailed in Table 1. The prevalence of medium and high-severity vulnerabilities underscores the substantial cybersecurity risk within these institutions.

### 4.2. Institution Characteristics

Examining the institutional records and characteristics by comparing them to their vulnerability exposure, we identified limited features correlating with the institution's number and type of vulnerabilities. The most critical institutional factors studied in this research are the institution's total yearly expenses and institution size (measured as full-time enrollment), which had an overall positive correlation. One such trend is that colleges with more students have more services running on average, as shown in Figure 2.b. Further, Figure 2.b also highlights that largerenrollment universities have more services running. This further followed that larger-enrollment universities

would have a significant increase in vulnerabilities, as shown in Figure 2.c. Figures 3.a, 3.b, and 3.c, indicate the relevant trends related to vulnerability scores and total expenses. Our analysis found no significant trends relating to other institutional characteristics (e.g., tuition costs, public vs. private, urban vs. rural location.)

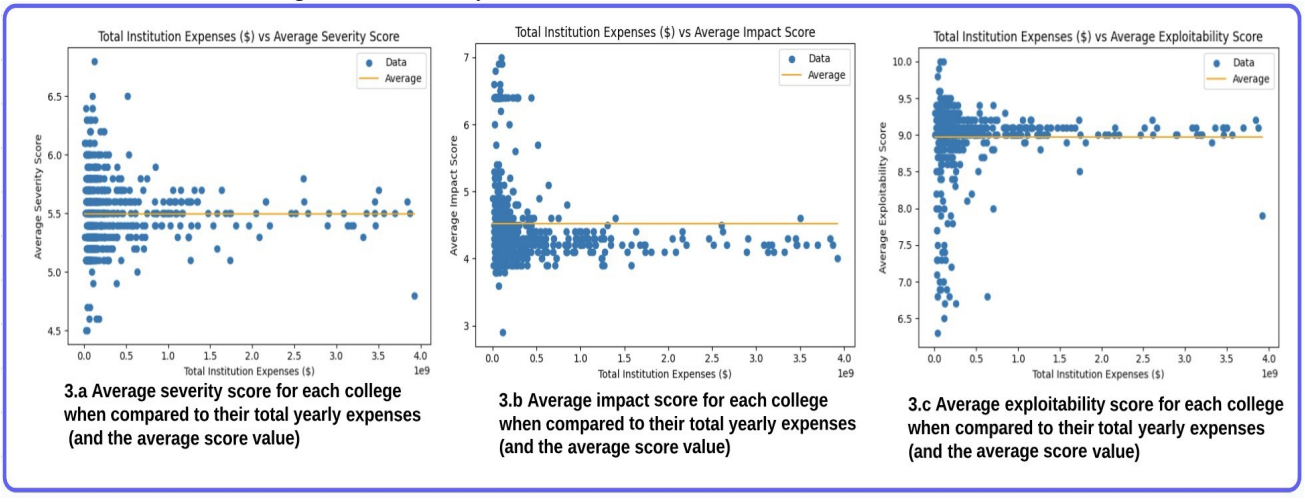



**2.a Classification of products by number of vulnerabilities according to Shodan analysis**

**2.b Number of average services for colleges of a given an institution's number of students**

**2.c Number of vulnerabilities plotted against the number of services at each college, with a linear fit trend-line showing the positive correlation**

Fig. 2. Vulnerability and Services Statistics with Institutional Size




**3.a Average severity score for each college when compared to their total yearly expenses (and the average score value)**

**3.b Average impact score for each college when compared to their total yearly expenses (and the average score value)**

**3.c Average exploitability score for each college when compared to their total yearly expenses (and the average score value)**

Fig. 3. Average severity, impact and exploitability score for each college when compared to their total yearly expenses

## 5.Discussion

In Figure 4.a, vulnerability age is crucial in university network security. However, there are still many vulnerabilities from earlier years, with 2018 being the average. Figure 4. b shows that despite time, the number and severity of vulnerabilities have remained steady since 2014. Table 1 reveals that school network vulnerabilities have a medium impact but are highly exploitable. Smaller schools have fewer vulnerabilities but face more severe and exploitable threats, unlike larger institutions with more services. Figure 5 confirms an increasing number of vulnerabilities with consistent severity levels. Figures 2.a and 2.b show larger institutions have more services and vulnerabilities. Smaller schools have higher variances in vulnerability, severity, impact, and exploitability variance, as shown in Figures 3.a, 3.b, and 3. c. The values suggest smaller schools face fewer vulnerabilities on average but with more significant potential impact. Figure 6 highlights a consistent vulnerability range across colleges, with

Apache and SSH as common targets proportionate to institution size. Due to the results of navigating through Apache and SSH as targets, these values strongly align with the potential possibility of ransomware attacks in smaller schools.

## 6.Conclusion

In this work, we examined vulnerabilities in higher education networks through Shodan scans, revealing a substantial average of forty-seven vulnerabilities per service, mostly with high exploit potential but of lower-medium impact. We investigated how an institution's characteristics, such as enrolment and expenses, relate to the type and quantity of network vulnerabilities. We uncovered that large-enrolment schools tended to have more vulnerabilities, while smaller ones had fewer but potentially more impactful vulnerabilities. We concluded that ransomware groups have a significant opportunity to target smaller enrolment United States universities due to their exposure.

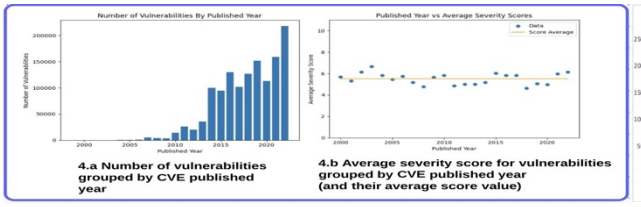



Fig. 4. Vulnerability statistics with Statistics

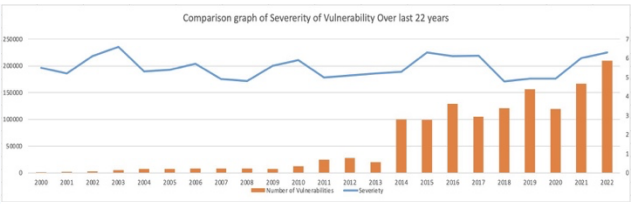



Fig. 5. Comparison of 20-year severity

| Service | under 1000 Vulnerability cou | 1000-4999 Vulnerability cou | 0-5times increase in size Percentage increase | 5000-9999 Vulnerability count | 5-10times increase in size Percentage increase | 10000-19999 Vulnerability count | 10-20times increase in size Percentage increase | above 20000 Vulnerability count | >20 times increase Percentage increase |
|---|---|---|---|---|---|---|---|---|---|
| NA | 20.8 | 34 | | 51 | | 76.5 | | 187 | |
| Apache Httpd | 12.3 | 20.5 | | 30.75 | | 46.13 | | 112.75 | |
| Microsoft IIS Httpd | 4.92 | 8.2 | | 12.3 | | 18.45 | | 45.1 | |
| openssh | 3.24 | 5.4 | | 8.1 | | 12.15 | | 29.7 | |
| nginx | 3.24 | 5.4 | | 8.1 | | 12.15 | | 29.7 | |
| Microsoft HTTP API httpd | 2.88 | 4.8 | | 7.2 | | 10.8 | | 26.4 | |
| Big IP | 2.46 | 4.1 | | 6.15 | | 9.23 | | 22.55 | |
| Cisco system | 1.2 | 2 | | 3 | | 4.5 | | 11 | |
| ntpd | 0.84 | 1.4 | | 2.1 | | 3.15 | | 7.7 | |
| other | 8.22 | 13.7 | 67 times increase | 20.55 | 150 times | 30.83 | 267 times | 75.35 | 800 times |

Fig. 6. Average range of vulnerabilities per product on all institution size

## 7.References

[1] Shodan [Internet]. (2023) [cited : 2023Dec18] Available from: https://www.shodan.io/
[2] Arin. [Internet]. (2023) [cited : 2023Dec21] Available from: https://whois.arin.net/
[3] Integrated Postsecondary Education Data System. [Internet]. (2023) [cited : 2023Dec18] Available from :https://nces.ed.gov/ipeds
[4] National Vulnerability Database. [Internet]. (2023) [cited : 2023Dec19] Available from: https://nvd.nist.gov/
[5] NESSUS. . [Internet]. (2023) [cited : 2023Dec11] Available from: https://www.tenable.com/products/nessus
[6] Nicholas Bogel-burroughs. (2019). [cited : 2023Dec18] Hackers' latest target: School districts. Available from :https://www.nytimes.com/2019/07/28/us/hacker-school-cybersecurity.html
[8] Augustín Chancusi, Paúl Diestra, and Damián Nicolalde. (2020.) Vulnerability analysis of the exposed public IPS in a higher education institution. 2020 the10th International Conference on Communication and Network Security
[9] Kate Gibson. (2022). ) [cited : 2023Dec19] Ransomware attack shutters 157-year-old Lincoln College Available from: https://www.cbsnews.com/news/lincoln-college-closes-ransomware-hackers-illinois/
[10] Emsisoft Malware Lab. (2022). [cited : 2023Dec19] The state of Ransomware in the US: Report and statistics 2021. Available from: https://www.emsisoft.com/en/blog/40813/the-state-of-ransomware-in-the-us-report-andstatistics-2021/
[11] Christopher Harrell, Mark Patton, and Sagar Samtani. 2018. Vulnerability Assessment, Remediation, and Automated Reporting: Case Studies of Higher Education
[12]Umesh Kumar, Chanchala Joshi, and Neha Gaud. 2016. Measurement of Security Dangers in University Network. International Journal of Computer Applications