XCAMPUS: EARLY WARNING FOR CAMPUS CYBER RISKS

Elvis Mwiseneza | Faculty Coauthor: Marlon Lagulos
School of technology and Innovation, Husson University |
mwisenezae1@mail.husson.edu

ABSTRACT

XCampus is a simple cybersecurity risk monitor for college networks. Its goal is to warn users when cyber risk is starting to rise before a major attack happens. Instead of only detecting attacks after they begin, the system looks for warning signs in network behavior. These signs include failed login attempts, suspicious IP access, unusual file downloads, new device connections, phishing related activity, and privileged access attempts.

To protect privacy, the project used realistic synthetic campus data instead of real student or staff records. The raw events were grouped by department and system, then turned into useful features such as after hours activity, outside access patterns, and sudden spikes in behavior. The project tested three methods: a machine learning model, an anomaly detection model, and a rule based method.

The final system gives each department or system a risk level of low, medium, or high and displays the results in a dashboard. In the latest test, the machine learning model performed best overall, with an F1 score of 0.73 and a ROC AUC of 0.84. These results show that behavior based monitoring can help universities notice cyber risk earlier, improve planning, and move from reacting to attacks toward preventing them.

BACKGROUND

- Colleges face phishing, ransomware, account takeover, and insider misuse.
- Many security tools are reactive and alert only after damage begins.
- Campus teams need a low cost way to see where cyber risk is rising.

Research questions:

1. Can behavior based machine learning predict future compromise risk?
2. Which event types are most linked to higher risk?
3. How does anomaly detection compare with rule based detection?

METHODS

- The project created 42,579 sample cybersecurity events from 7 campus departments and 14 systems.
- It tracked events such as failed logins, suspicious IP access, unusual downloads, new devices, phishing signs, and privileged access attempts.
- These events were grouped into 2,520 daily records by department and system.
- The system then measured useful patterns like after hours activity, outside network access, serious events, weekly averages, and sudden spikes.
- Three methods were compared: machine learning, anomaly detection, and rule based detection.
- The results were combined into one final risk score and shown on a Streamlit dashboard as low, medium, or high risk.

RESULTS AND DASHBOARD HIGHLIGHTS

Performance summary

Precision tells us if an alert is believable, recall tells us if we are missing threats, F1 balances both, and ROC AUC shows how well the model separates risky systems from safe ones.

- Predictive ML: Accuracy 0.77 | Precision 0.76 | Recall 0.70 | F1 0.73 | ROC AUC 0.84

Most Important Drivers

- The most helpful warning signs were the total number of events, number of devices, number of users, serious security events, outside IP activity, and sudden changes in behavior.

Latest Findings

- The departments with the highest risk were Finance, Library, Student Services, and Research Lab.
- The systems with the highest risk were research data vault, library catalog, payroll db, and housing api.

What the Dashboard Shows

- The dashboard displays summary cards, model performance results, department risk levels, system risk levels, changes in risk over time, and the main signals causing the risk score to rise.

CONCLUSIONS

- This project shows that network behavior can give early warning signs before a cyberattack becomes serious.
- The final risk score combines different methods, which makes the system both smart and easy to understand.
- Even though the project used synthetic data, it still shows how schools can track and manage cyber risk in a practical way.
- A good next step would be testing the system with real campus security logs from tools like Zeek, Splunk, or ELK.

REFERENCES

- Harvard University computer science and machine learning course.
- Breiman L. Random Forests. Machine Learning. 2001.
- Liu FT, Ting KM, Zhou ZH. Isolation Forest. ICDM. 2008.
- Scikit learn Developers. Scikit learn documentation.
- CISA. Phishing guidance and cybersecurity best practices.

ACKNOWLEDGEMENTS

- Harvard University
- Husson University Business School
- Faculty coauthor and project advisor: [Marlon Lagulos]
- Open source tools used: Python, Github, Pandas, Scikit learn, and Streamlit
- No human subjects were used; all data in this prototype were synthetic.