From Probability to Protection: Poisson Inequalities for Machine Learning in Cybersecurity

Dr. S. Spector E. Ibokete

Department of Mathematics & Computer Science
Canisius University
Buffalo, NY, USA

Supported by the Koessler Summer Fellowship

The Problem: Alert Fatigue in Cybersecurity

The Crisis

74% of security alerts are IGNORED

Security analysts are overwhelmed by false alarms

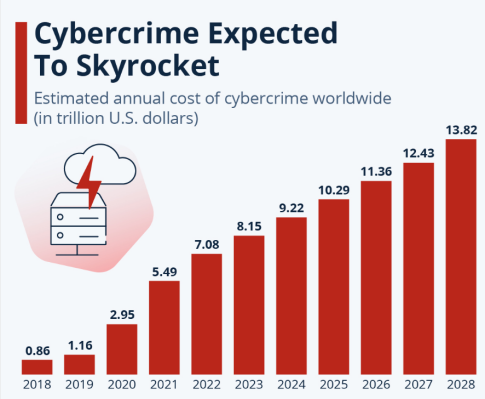
Annual Security Costs

$5.3M

Per organization

Why Current ML Systems Fail

- × No mathematical guarantees
- × Cannot explain decisions
- × Operate as "black boxes"



Our Solution: Mathematical Certainty

Network Traffic Follows Patterns

Like customers arriving at a bank or calls at a call center

We use **Poisson distributions** to model normal behavior

Our Innovation: Khinchine-type Inequalities

The Mathematical Guarantee:

$$\mathbb{E}\left[\left|\sum_{i=1}^{N} a_i X_i\right|^q\right] \leq T_q\left(\sum_{i=1}^{N} \lambda_i\right) \|a\|_\infty^q$$

- ✓ Provides **certified confidence intervals**
- ✓ Mathematical proof: "This is anomalous with 99% confidence"
- ✓ Not a guess—a **mathematical guarantee**

Real-World Example: Log4Shell Attack (December 2021)

What is Log4Shell?

A "hidden door" vulnerability in logging software used by **millions of companies**

How the Attack Works

1. Hackers send special requests
2. System performs JNDI lookup
3. Downloads malicious code
4. System compromised

Our Detection

Normal Behavior:

$\lambda = 0.3$ lookups/min

During Attack:

15 lookups/min

$$P < 10^{-15}$$

1 in a quadrillion!

Mathematical Certainty

Not a guess!

Implementation & Results

Real-Time Detection System

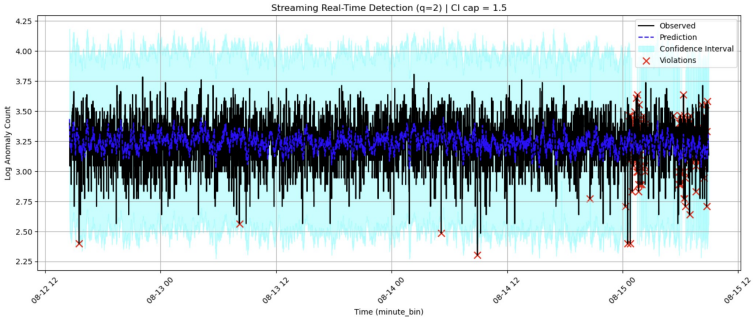
- Processes traffic in **real-time**
- Works with **encrypted traffic**
- Runs on **standard hardware**

Detection Accuracy

89–94%
Competitive with ML methods

Confidence Guarantee

95%+
Mathematical proof included



Tested on Real Attack Data

650,570	**153 days**	**7 types**
network flows	monitoring period	attack categories

Student Opportunities & Impact

How Students Get Involved

Data Analysis Work with real cybersecurity datasets

Mathematical Modeling Apply probability theory to real problems

Software Development Build detection algorithms in Python

Skills Students Gain

- Machine Learning
- Network Security
- Statistical Analysis
- Research Methods
- Python Programming
- Academic Publishing

Real Impact

Conference Submission

Open Source Code

Your work protects real systems!