

the analysis of user behavior in enterprise networks

Prepared by

ALI MMADI, PhD student - Midocean University



Research Skills (1)

Task 7

ALI MMADI

PhD student - Midocean University

www.midocean.ae

A Behavioral Approach to Insider Threat Detection Using User Activity Logs

INTRODUCTION

In this video, I will present to you a concrete example drawn from scientific research in the field of cybersecurity. The aim is to show you how a study defined its study community and its sample, what type of sampling was used, and how the researcher calculated the sample size. I will only provide the elements reported in the research, without critiquing them.



Presentation of the chosen research

The study I chose is titled:

"A Behavioral Approach to Insider Threat Detection Using User Activity Logs" published in 2021 by the research team at Carnegie Mellon University. This research aimed to analyze user behaviors within corporate computer networks in order to detect suspicious or malicious behaviors related to insider threats. To do this, the researchers relied on real data collected from a large organization that granted access to its login and user activity logs.

Study community

The study community consisted of all employees using the company's computer systems, approximately 2,800 people. This population was diverse, including executives, technicians, administrative staff, and external consultants. This community represented an interesting real case for the study of cybersecurity behavior, as it included different levels of access and activities on the network.

Selected sample and type of sampling

For feasibility reasons, the researchers did not study the entire community. They used stratified random sampling.

- This means they divided the population into groups (or strata) according to professional roles (for example: administrators, standard users, consultants), and then randomly selected a proportional number of individuals from each group. This ensured that each user category was well represented in the final sample.

Sample size and equation used

To calculate the sample size, researchers relied on the classic formula for calculating a random sample in a finite population:

$$n = \frac{N \cdot Z^2 \cdot p(1 - p)}{(N - 1) \cdot e^2 + Z^2 \cdot p(1 - p)}$$

With:

N = size of the population (here 2800)

Z = value of the normal distribution for a confidence level of 95% (Z = 1.96) pp = estimated proportion (often 0.5 for a maximum estimate)

e = tolerated margin of error (here 5%, or 0.05)

By substituting the values, they obtained a representative sample of about 340 individuals, distributed according to the previously defined strata.

CONCLUSION

In summary, this study demonstrates a rigorous application of the scientific methodology in the field of cybersecurity. It precisely defined its study community, chose a random stratified sampling method, and calculated the sample size using a well-known formula in statistics.

Thank you for your attention and see you soon for another analysis of scientific research.

Main research references

- 1) The CERT Insider Threat Dataset r4.2, Community Emergency Response Team, Carnegie Mellon University — used in the study on internal threat detection.[arXiv+12MDPI+12ResearchGate+12](#)
- 2) Noever, D. A. "Classifier Suites for Insider Threat Detection" — comparative evaluation of several algorithms on the CERT dataset, with results exceeding 98% accuracy.[arXiv+1MDPI+1](#)
- 3) Gayathri et al. (2022). Approach combining GAN and supervised learning on the CERT dataset to enrich the rare cases of internal attackers.[MDPI](#)



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY



Thank You

