DOI:10.1145/3689596

Peter G. Neumann and Ulf Lindqvist

Inside Risks
The Future of Misuse Detection

From lessons learned to new directions.

Misuse detection is a term loosely encompassing real-time detection of attacks, insider misuse, malware, and other security violations. Its difficulties are widespread, including inherent incompleteness of methods, lack of certainty and assurance in the real-time results (for example, too many false positives and false negatives), dynamic changes in profiles over time (although updating became a business driver), and opportunities for counterproductive training for the insider profile-based approach.

This topic has a long and broad history, with many publications occurring in the 1980s—although there had been various seeds planted earlier (for example, Anderson[2]). Note that we prefer the term "misuse detection"[7] to the narrower and misleading term "intrusion detection," which is often used confusingly in a broader sense that is not limited to intrusions.

To illustrate the evolution of work in this area, we describe a sequence of interrelated projects at the SRI International Computer Science Lab, not to hype them but rather as representative examples of the body of earlier work in this field. This work began in 1983 with our profile-based insider misuse detection system using statistical analysis.[5] We then incorporated an early rule-based expert system P-BEST[8,15] into SRI's Intrusion Detection Expert System (IDES).[3,4,9] We next combined the first two projects into the Next-generation Intrusion Detection System (NIDES) that used both rule-based and profile-based detection.[1] A few years later came EMERALD: Event Monitoring Enabling Responses to Anomalous Live Disturbances,[10,12] which developed and composed a distributed scalable tool suite for deploying real-time detection, analysis, correlation, and response technologies in a hierarchical fashion across large networks. The EMERALD technology was licensed to several security industry leaders and overcame some of the problems with the earlier work.

This unified collection of efforts was similar to work by other research teams that also encountered many of these difficulties in approximately the same time period.

We also note that misuse detection encounters even more difficulties in trying to identify and eliminate or otherwise address intentional malware—whether in real time or in analyses of source code and object code. In addition, it should be clear to readers of *Communications* Inside Risks columns that we have a huge shortage of trustworthiness: We should not trust the existing hardware because of inherent vulnerabilities. Today's bulky operating systems get frequent critical updates because new flaws keep surfacing, implying that we should not trust the total systems and networks—especially when used for applications (for example, with AI) that are life-critical or mission-critical.[11] Almost nothing today

is trustworthy enough to withstand misuse and attacks.

The risks resulting from attacks and misuse have escalated dramatically in recent years, with ransomware, targeted phishing attacks, social engineering including disinformation and media exploitation, supply-chain attacks, hardware tampering, and much more. However, these burgeoning problems lie outside the realm of misuse detection, but are nevertheless worth mentioning as part of the overall trustworthiness problem.

A retrospective conclusion is that misuse detection was generally a very difficult problem in the previous century; however, addressing the overall threats is even more challenging today because of the escalation in potential risks—not just to the computer systems and networks, but also all the other issues such as human factors.

What Can We Expect in the Future?

The techniques and tools for misuse detection and remediation in the past seem to be intrinsically limited. Thus, we might want to seek some radical new approaches. This thought leads us to the question: Would any elements of AI help us today or in the future?

There was a prescient 1989 report by John Rushby and Alan Whitehurst[13] on the potential use of formal verification of AI systems to increase the assurance of such systems. In retrospect, that report has become extremely relevant today in light of the enormous lack of assurance in today's AI feeding frenzy in commercial AI.[14]

Considering the assurance issues, we believe AI researchers have the potential to save AI from the hype, as long as they increasingly place substantial emphasis on evidence-based research. One recent example is given by Jha, Rushby, and Shankar.[6] Other efforts are also ongoing that train chatbots properly—and then independently assess (perhaps even formally prove) the correctness of the results. In stark contrast with the burgeoning research on AI assurance, the emerging commercial AI today exhibits very little in terms of trustworthiness and other guarantees with respect to the quality of the output.

Various AI and machine-learning techniques have been applied to misuse detection since its conception, with varying degrees of success in limited areas but without addressing the fundamental problems. It appears that the inherent difficulties of misuse detection would still remain fundamentally unwieldy, even if a surge of evidence-based AI were to take place. Complex problems rarely have easy solutions.

The techniques and tools for misuse detection and remediation in the past seem to be intrinsically limited.

There are areas related to misuse detection where trustworthy AI would have the potential to greatly augment human defenders, and where some tasks could be automated and performed in real time, while others are of a more interactive and investigative nature. Anomaly detection applied to event data from hosts and networks, malware detection, alert prioritization, detection of spam and phishing messages, and transaction fraud detection are all areas where AI is currently used and would benefit from increased trustworthiness. Interactive chatbots enabled by large language models (LLMs) have the potential to boost the performance of human investigators if they could be trusted to summarize complex events, allow the use of natural-language dialog queries into large datasets, suggest and guide response actions, and generate incident reports.

The rapid development of untrustworthy commercial AI tools will also help attackers become more productive and effective. For example, AI can help attackers evade detection tools by improving their ability to learn and mimic legitimate behavior. The big question is: Will AI help defenders improve faster and better than attackers, or will AI simply amplify the asymmetric imbalance that already exists—which typically favors the attackers?

In short, some earlier work in the previous century was already a harbinger for many of the problems commercial AI is exhibiting today, which suggests new efforts with evidence-based trustworthiness are going to be essential in the future. New research is needed (possibly less stringent than formal analysis) to provide higher assurance. Of course, in hindsight, the best strategy would be to have built everything much more carefully in the first place. C

References

1. Anderson, D., Frivold, T., and Valdes, A. *Next-Generation Intrusion-Detection Expert System (NIDES)*. Tech. Rep., Computer Science Laboratory, SRI International, Menlo Park, CA, SRI-CSL-95-07, (May 1995).
2. Anderson, J.P. *Computer Security Technology Planning Study*. Tech. Rep. ESD-TR-73-51, ESD/AFSC, Hanscom AFB, Bedford, MA, (Oct. 1972).
3. Denning, D.E. et al. An intrusion-detection model. *IEEE Trans. on Software Engineering 13*, 2 (Feb. 1987).
4. Denning, D.E. et al. *A Prototype IDES: A Real-Time Intrusion-Detection Expert System*. Tech. Rep., Computer Science Laboratory, SRI International, Menlo Park, CA (1987).
5. Javitz, H.S. et al. *Analytical Techniques Development for a Statistical Intrusion-Detection System (SIDS) Based on Accounting Records*. Tech. Rep., SRI International, Menlo Park, CA (July 1986).
6. Jha, S., Rushby, J., and Shankar, N. Model-centered assurance for automonous systems. In *Computer Safety, Reliability, and Security, vol. 12234 of Springer Lecture Notes on Computer Security* (2023); https://bit.ly/3T6PpES
7. Lindqvist, U. On the fundamentals of analysis and detection of computer misuse. Ph.D. dissertation, Department of Computer Engineering, Chalmers University of Technology, (1999).
8. Lindqvist, U. and Porras, P. Detecting computer and network misuse through the Production-Based Expert System Toolset (P-BEST). In *Proceedings of the 1999 Symp. on Security and Privacy* (Oakland, CA, May 1999), IEEE Computer Society.
9. Lunt, T. et al. *A Real-Time Intrusion-Detection Expert System (IDES)*. Tech. Rep., Computer Science Laboratory, SRI International, Menlo Park, CA (Feb. 28, 1992).
10. Neumann, P. and Porras, P. Experience with EMERALD to date. In *Proceedings of the First USENIX Workshop on Intrusion Detection and Network Monitoring* (Santa Clara, CA, Apr. 1999).
11. Neumann, P.G. Toward total-system trustworthiness. *Commun. ACM 65*, 6 (June 2022); https://bit.ly/3X5SuGp
12. Porras, P. and Neumann, P. EMERALD: Event monitoring enabling responses to anomalous live disturbances. In *Proceedings of the 19th National Computer Security Conf.* (Baltimore, MD, Oct. 1997).
13. Rushby, J. and Whitehurst, R.A. Formal verification of AI software. Final report, NASA contract 18226 (task5), Computer Science Laboratory, SRI International, Menlo Park, CA (Feb. 1989); https://bit.ly/3YboryE
14. Schneier, B. AI and Trust. Belfer Center for Science and International Affairs, Harvard Kennedy School (Nov. 27, 2023); https://bit.ly/3Xmsnw7
15. Sebring, M. et al. Expert system in intrusion detection: A case study. In *Proceedings of the 11th National Computer Security Conf.* (Baltimore, MD, Oct. 1988).

Peter G. Neumann (peter.neumann@sri.com) is chief scientist at the SRI International Computer Science Lab in Menlo Park, CA, USA, and creator/moderator of the ACM Risks Forum.

Ulf Lindqvist (ulf.lindqvist@sri.com) is senior technical director at the SRI International Computer Science Lab in San Luis Obispo, CA, USA.