

Paper to Practice

The Importance of Systems Thinking in Machine Learning for Cybersecurity

Scott Coull

February 26, 2024

About Me



Scott Coull

Cloud Security
Head of DS Research

More than 20 years at the intersection of Cybersecurity and Machine Learning.

Over that time, research interests have spanned multiple problems areas, including data privacy, network traffic analysis, malware analysis, security operations, and applied cryptography

Lead development of Format-Transforming Encryption, MalwareGuard, and many other projects that have successfully transitioned from research to practice and protect millions of users each day

Excited about exploring problems at the intersection of research and practice, particularly when research assumptions do not align well with practice

Machine Learning for Cybersecurity

Applying Machine Learning approaches to Cybersecurity problems (Sec+ML) was first proposed nearly 40 years ago by Dorothy Denning¹

Sec+ML has since grown into a pervasive force, with nearly every aspect of cybersecurity leveraging ML in some way to provide better outcomes for users

- Examples include: vulnerability detection, malware analysis, security event prioritization, and network traffic analysis

Critically, these **ML models almost never stand alone** and are often just one component of much larger interconnected systems that operate holistically

- Pitfalls and nuances of such systems studied by Sculley et al.²



[1] Denning, Dorothy E. "An intrusion-detection model." IEEE Transactions on software engineering 2 (1987): 222-232.

[2] Sculley, David, et al. "Machine learning: The high interest credit card of technical debt." SE4ML: Software Engineering for Machine Learning (2014).

If the **system** plays such a pivotal role in the **success of Sec+ML**, why don't we **focus** more on **studying it holistically**?

Objectives

Introduce the concept of systems thinking and how it applies to Sec+ML

Highlight how lack of systems thinking hampers transition of research into practice

Generalize lessons learned and suggest paths forward for the Sec+ML community

Disclaimers

All **views are my own and are based on my experience** at the intersection of research and practice in Sec+ML

This is not meant as a critique of any particular research direction/paper

Lots of **good work already started in this direction** in the context of adversarial ML, such as work by Apruzzese et al.³ and Grosse et al.⁴

View this as an opportunity to mature the relationship between research and practice in this critical area, which ultimately **increases impact**







[3] Apruzzese, G. et al. "Real Attackers Don't Compute Gradients": Bridging the Gap Between Adversarial ML Research and Practice. 2023 IEEE Conf. Secur. Trust. Mach. Learn. (SaTML), 339–364 (2023).
[4] Grosse, Kathrin, et al. "Machine learning security in industry: A quantitative survey." IEEE Transactions on Information Forensics and Security 18 (2023): 1749-1762.

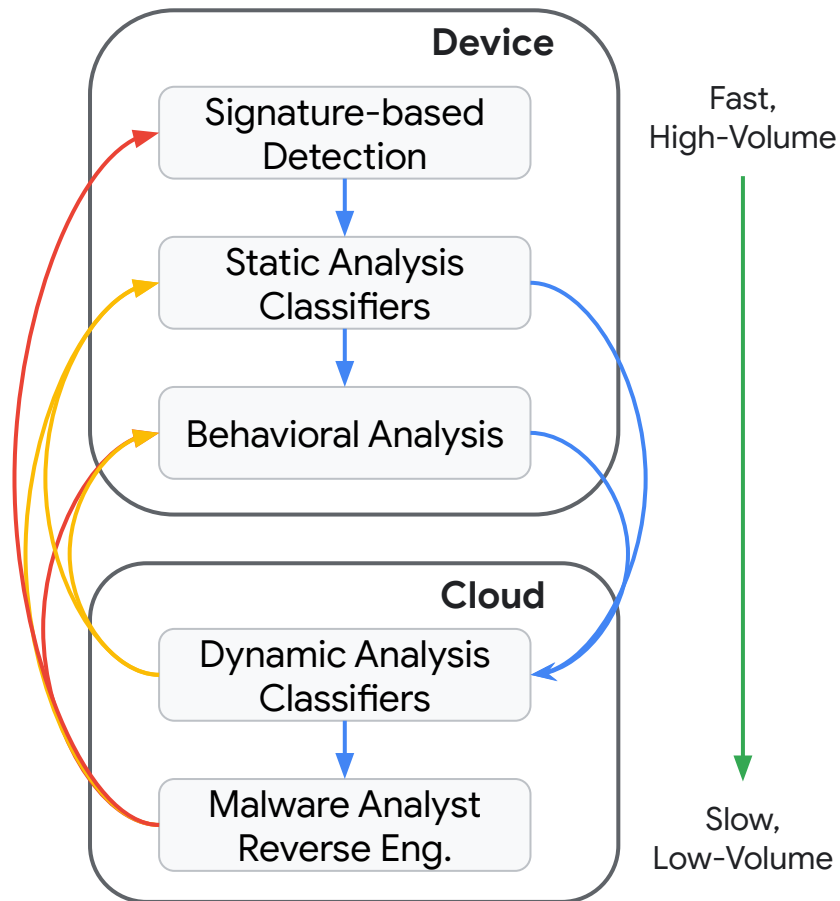
Systems Thinking

The behavior (and efficacy) of complex systems cannot be adequately captured by examining their constituent parts independently, instead we must **consider interactions that lead to (unexpected) emergent behaviors:**

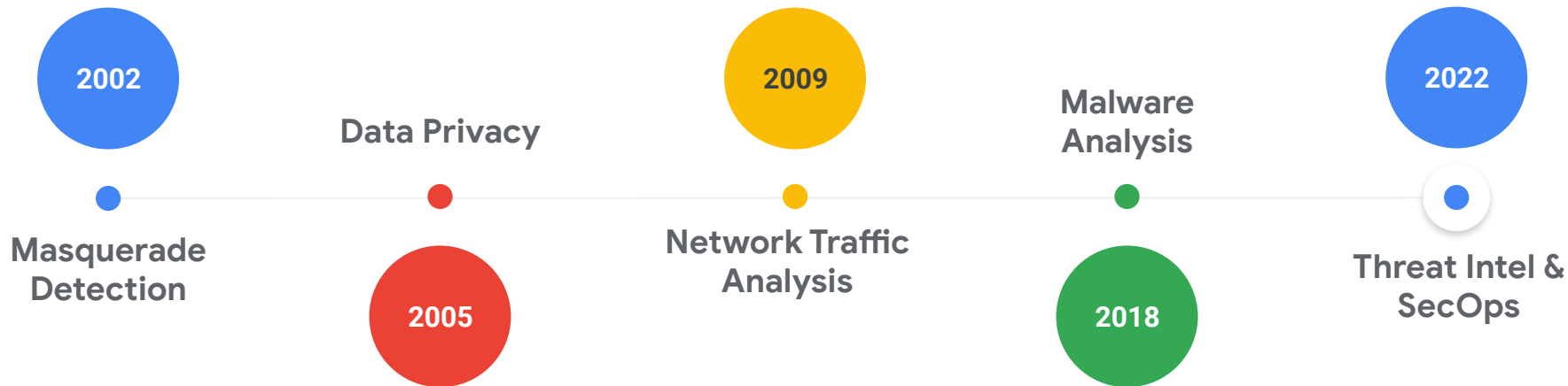
- Interconnectedness
- Dependency
- Feedback loops

Demonstrated in real-world malware classification system architectures:

-  Routing based on outcome
-  New signatures and rules
-  Updated labels and samples
-  Environment restrictions/requirements

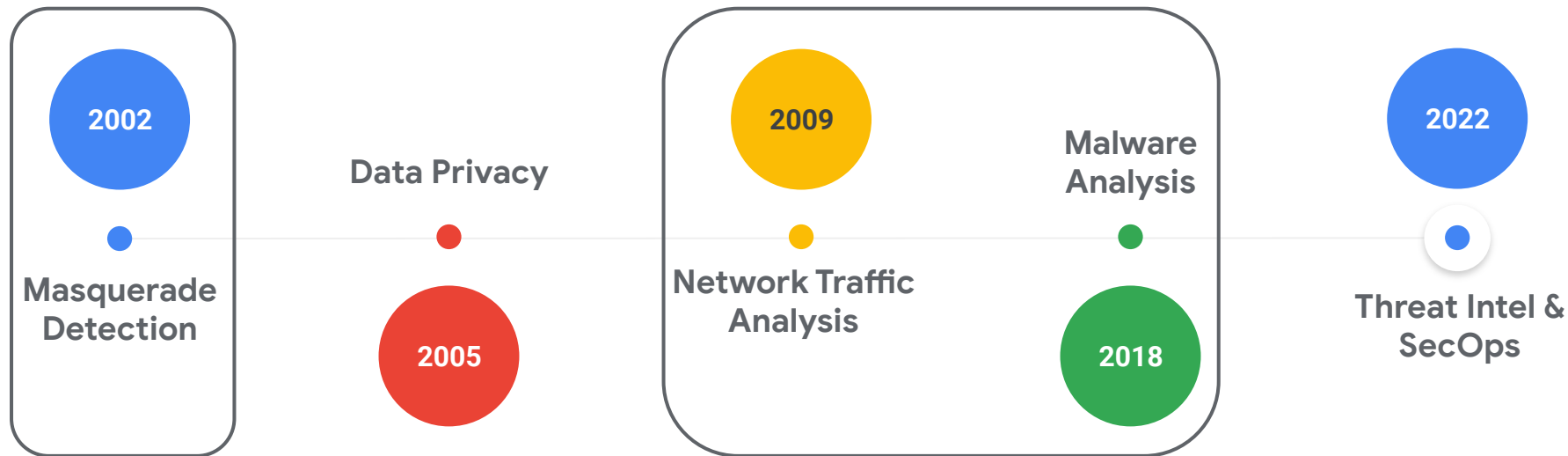


The Eras Tour of Sec+ML



Let's explore these concepts by looking at how **real-world systems concepts** have intersected with my research career

The Eras Tour of Sec+ML



Let's explore these concepts by looking at how **real-world systems concepts** have intersected with my research career

Masquerade Detection

The difficulty of studying systems-level problems was evident from the start – publishing usually **requires a recognized research problem** and **benchmarks to show progress**

Masquerade detection is the problem of identifying when a user account has been taken over or is being abused by an attacker – one of the core problems posed by Denning in 1987.

Dataset generated from live user command line usage to simulate masquerading published by Schonlau et al.⁵ was readily available and had multiple published benchmarks on the task

The **problem setup and dataset was unrealistic** in a number of ways but the ability to easily demonstrate **SOTA improvements over prior work made it 'easy' to publish**



[5] Schonlau, Matthias, et al. "Computer intrusion: Detecting masquerades." Statistical science (2001): 58-74.

Lesson #1: Momentum

- It is easier to publish on well-established problems
- Public datasets, even unrealistic ones, can significantly lower the bar to entry (and are difficult to stop once they take hold)
- Systems can be difficult to define, quantify, and share captured data about
- Systems-level research is therefore seen as a risk even when individual components have reached maturity

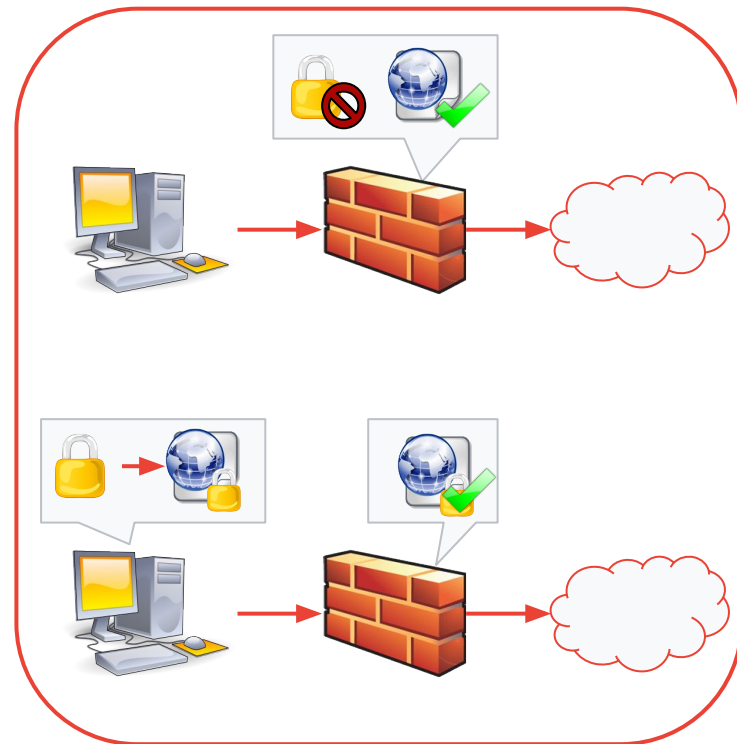
Network Traffic Analysis

Work on network traffic analysis spanned both evasion and detection problems:

- Methods to **infer information from the features of encrypted communications** other than its contents, such as message timing and size.
- Tools and techniques to **circumvent censorship by hiding the presence of encryption** so that it was difficult to detect and block.

Here, the core question always centered around what is realistic for an adversary:

- What is an **acceptable false positive rate** for detection in real network settings?
- What **traffic mix or other real-world artifacts** may be present in the data?
- What are the **capabilities of adversaries** in monitoring multi-gigabit links?



DPI boxes are very complicated and expensive.

They do **much more to classify traffic** than what Squid or Wireshark would do. The real test is whether an adversary with **high-end DPI equipment** ...

- USENIX HotSec Reviewer

In fact, there are significant environmental restrictions when monitoring multi-gigabit links that limit the amount of state and depth of analysis that can be performed.

Later analysis of an enterprise-grade DPI platform showed it was **less** capable than some open source alternatives



DPI boxes are very complicated and expensive.

They do **much more to classify traffic** than what Squid or Wireshark would do. The real test is whether an adversary with **high-end DPI equipment** ...

- USENIX HotSec Reviewer

... **low false positive rate** for conservative values is less encouraging, since that means an interested regime would be **all too happy to let 3% of queries fail** in order to **block this attack**.

- IEEE S&P (Oakland) Reviewer

Failure to understand the realities of the base-rate fallacy in a large-scale network traffic analysis setting.

3% FP rate would amount to millions or billions of failed network connections.

... **low false positive rate** for conservative values is less encouraging, since that means an interested regime would be **all too happy to let 3% of queries fail** in order to **block this attack**.

- IEEE S&P (Oakland) Reviewer

Lesson #2: Environment

- Some research problems are inextricably linked with the constraints and restrictions of real-world environments
- Lacking any authoritative source, it is difficult to determine what is a reasonable expectation for performance
- Performing research with bad assumptions on restrictions leads to technologies with little chance of transitioning to practice
- Research that takes advantage of or incorporates those restrictions into the solution design may be unfairly judged

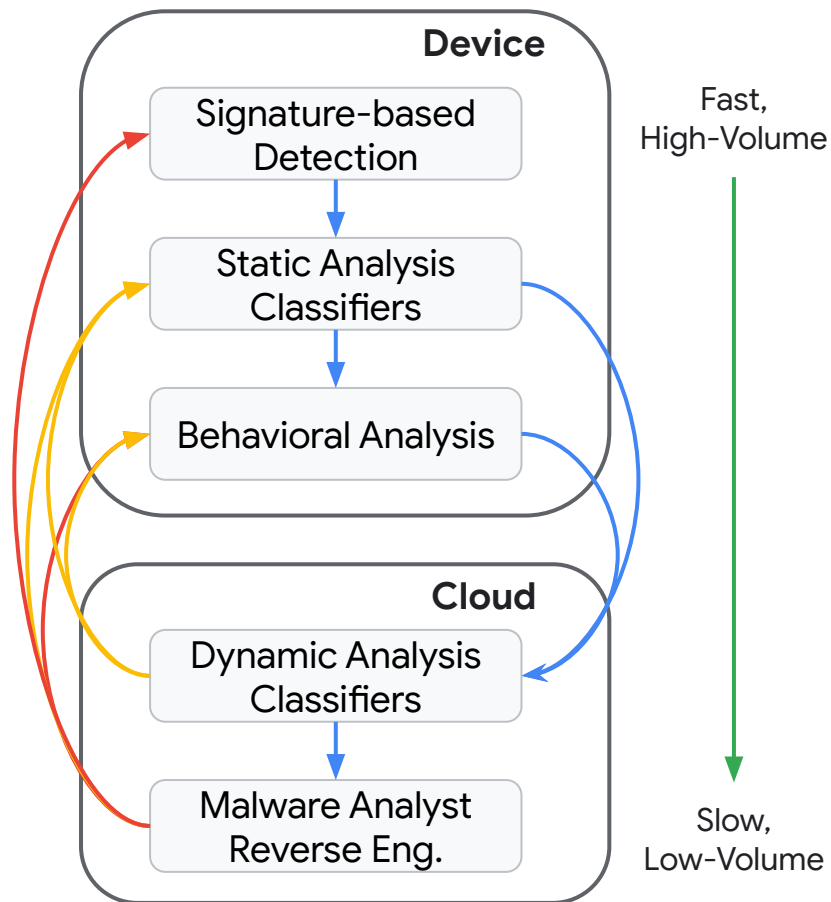
Malware Analysis

Explored both new classifiers and detection methods, as well as evasion and poisoning attacks on malware models

- **Malware detection and family classification** based on both static and dynamic analysis features
- **Adversarial example** generation and **clean-label backdoor poisoning** on static analysis-based malware models

To understand the impact of attacks or value of new detection techniques, we need to put them in context they operate in

For instance, to determine if a poisoning attack is viable in practice, we need to **understand the source of training data and labels**



The evaluation is **limited** to **static PE features**.

As a consequence, the **real-world impact** of this work remains **unclear**.

- USENIX Security Reviewer

Lack of understanding of the malware analysis pipeline and the reasons for implementing fast static analysis-based detections.

Assumption that because it is weaker than dynamic analysis methods it must not be used or valuable in practice.

The evaluation is **limited** to **static PE features**.

As a consequence, the **real-world impact** of this work remains **unclear**.

- USENIX Security Reviewer

Lesson #3: Architecture

- The true value of a detection method or practicality of an attack can only be determined in the context of its broader system
- Although a technique may be strictly weaker in an absolute sense, it may play an important role in a holistic defense-in-depth strategy
- Like environment, without guidance on what real-world architectures look like, we are left with questionable impact

Lessons Learned

Lesson #1: Momentum

- Difficult to motivate work on systems problems that are unfamiliar to the research community
- Lack of datasets, benchmarks, and clear system-level problem statements add to the challenge

Lesson #2: Environment

- Real-world impact is often mitigated by environmental factors that may be unknown to researchers
- Exploration of unfruitful research directions as well as suppression of viable approaches

Lesson #3: Architecture

- Strength or weakness of an attack or detection method is not absolute, but dependent on system-level context
- Focus on overly-specific areas with minor impact when holistic design may reap greater rewards

Where From Here?

The key, of course, is encouraging more collaboration and information sharing between academia and industry, but how do we do that?

The goal should not be to know the details of every system implementation, and instead to build a shared understanding.

Encouraging industry engagement with academia:

- More industry experts on academic program committees
- Publish reference architectures in white papers and blog posts

Encourage academic engagement with industry:

- Internships and sabbaticals in industry product groups (not just labs!)
- Technology transfer and R&D collaboration programs

A conference venue specifically designed for systems-level issues in Sec+ML!

- Real World Crypto Symposium⁶ is an excellent example in applied crypto



[6] <https://rwc.iacr.org/>



Thank you.