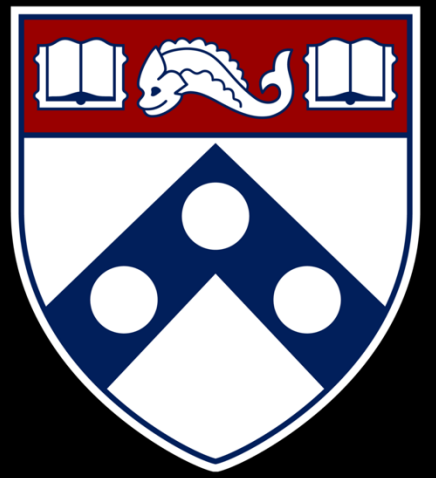
Secure Systems Engineering and Management

A Data-driven Approach



Cybersecurity Economics and Silver Bullets

Michael Hicks

UPenn CIS 7000-003, Spring 2026

Readings

Why Information Security is Hard
– An Economic Perspective

Ross Anderson

University of Cambridge Computer Laboratory,
JJ Thomson Avenue, Cambridge CB3 0FD, UK
Ross.Anderson@cl.cam.ac.uk

Abstract

According to one common view, information security comes down to technical measures. Given better access control policy models, formal proofs of cryptographic protocols, approved firewalls, better ways of detecting intrusions and malicious code, and better tools for system evaluation and assurance, the problems can be solved.

In this note, I put forward a contrary view: information insecurity is at least as much due to perverse incentives. Many of the problems can be explained more clearly and convincingly using the language of microeconomics: network externalities, asymmetric information, moral hazard, adverse selection, liability dumping and the tragedy of the commons.

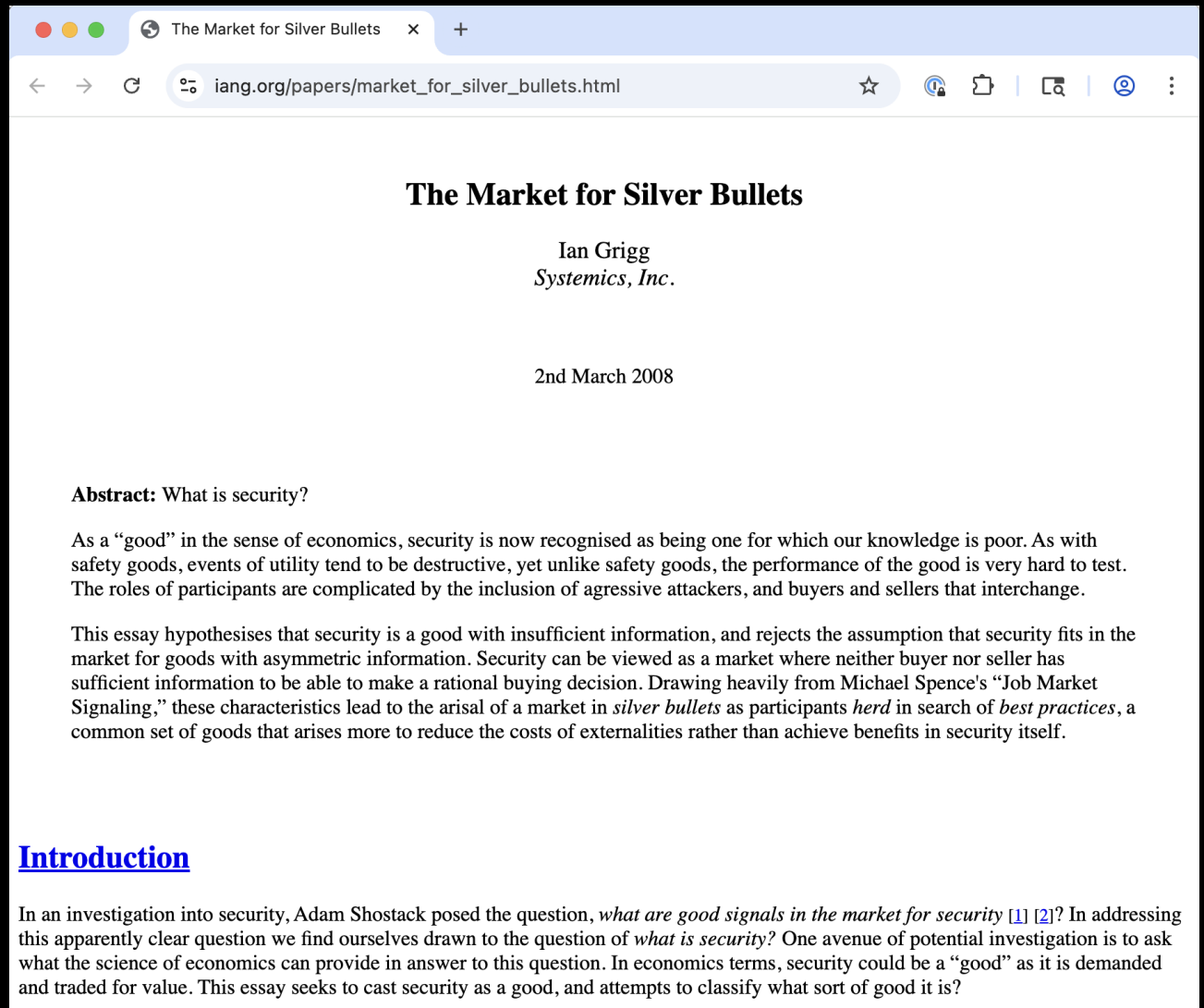
1 Introduction

In a survey of fraud against autoteller machines [4], it was found that patterns of fraud depended on who was liable for them. In the USA, if a customer disputed a transaction, the onus was on the bank to prove that the customer was mistaken or lying; this gave US banks a motive to protect their systems properly. But in Britain, Norway and the Netherlands, the burden

risk of forged signatures from the bank that relies on the signature (and that built the system) to the person alleged to have made the signature. Common Criteria evaluations are not made by the relying party, as Orange Book evaluations were, but by a commercial facility paid by the vendor. In general, where the party who is in a position to protect a system is not the party who would suffer the results of security failure, then problems may be expected.

A different kind of incentive failure surfaced in early 2000, with distributed denial of service attacks against a number of high-profile web sites. These exploit a number of subverted machines to launch a large coordinated packet flood at a target. Since many of them flood the victim at the same time, the traffic is more than the target can cope with, and because it comes from many different sources, it can be very difficult to stop [7]. Varian pointed out that this was also a case of incentive failure [20]. While individual computer users might be happy to spend $100 on anti-virus software to protect themselves against attack, they are unlikely to spend even $1 on software to prevent their machines being used to attack Amazon or Microsoft.

This is an example of what economists refer to as the 'Tragedy of the Commons' [15]. If a hundred peas-



The Market for Silver Bullets

Ian Grigg
Systemics, Inc.

2nd March 2008

Abstract: What is security?

As a "good" in the sense of economics, security is now recognised as being one for which our knowledge is poor. As with safety goods, events of utility tend to be destructive, yet unlike safety goods, the performance of the good is very hard to test. The roles of participants are complicated by the inclusion of agressive attackers, and buyers and sellers that interchange.

This essay hypothesises that security is a good with insufficient information, and rejects the assumption that security fits in the market for goods with asymmetric information. Security can be viewed as a market where neither buyer nor seller has sufficient information to be able to make a rational buying decision. Drawing heavily from Michael Spence's "Job Market Signaling," these characteristics lead to the arisal of a market in *silver bullets* as participants *herd* in search of *best practices*, a common set of goods that arises more to reduce the costs of externalities rather than achieve benefits in security itself.

Introduction

In an investigation into security, Adam Shostack posed the question, *what are good signals in the market for security* [1] [2]? In addressing this apparently clear question we find ourselves drawn to the question of *what is security?* One avenue of potential investigation is to ask what the science of economics can provide in answer to this question. In economics terms, security could be a "good" as it is demanded and traded for value. This essay seeks to cast security as a good, and attempts to classify what sort of good it is?

Readings

Essentially **founded the field of security economics**

Why Information Security is Hard
– An Economic Perspective

Ross Anderson

University of Cambridge Computer Laboratory,
JJ Thomson Avenue, Cambridge CB3 0FD, UK
Ross.Anderson@cl.cam.ac.uk

Abstract

According to one common view, information security comes down to technical measures. Given better access control policy models, formal proofs of cryptographic protocols, approved firewalls, better ways of detecting intrusions and malicious code, and better tools for system evaluation and assurance, the problems can be solved.

In this note, I put forward a contrary view: information insecurity is at least as much due to perverse incentives. Many of the problems can be explained more clearly and convincingly using the language of microeconomics: network externalities, asymmetric information, moral hazard, adverse selection, liability dumping and the tragedy of the commons.

1 Introduction

In a survey of fraud against autoteller machines [4], it was found that patterns of fraud depended on who was liable for them. In the USA, if a customer disputed a transaction, the onus was on the bank to prove that the customer was mistaken or lying; this gave US banks a motive to protect their systems properly. But in Britain, Norway and the Netherlands, the burden

risk of forged signatures from the bank that relies on the signature (and that built the system) to the person alleged to have made the signature. Common Criteria evaluations are not made by the relying party, as Orange Book evaluations were, but by a commercial facility paid by the vendor. In general, where the party who is in a position to protect a system is not the party who would suffer the results of security failure, then problems may be expected.

A different kind of incentive failure surfaced in early 2000, with distributed denial of service attacks against a number of high-profile web sites. These exploit a number of subverted machines to launch a large coordinated packet flood at a target. Since many of them flood the victim at the same time, the traffic is more than the target can cope with, and because it comes from many different sources, it can be very difficult to stop [7]. Varian pointed out that this was also a case of incentive failure [20]. While individual computer users might be happy to spend $100 on anti-virus software to protect themselves against attack, they are unlikely to spend even $1 on software to prevent their machines being used to attack Amazon or Microsoft.

This is an example of what economists refer to as the 'Tragedy of the Commons' [15]. If a hundred peas-

Attack vs. Defense Asymmetry

Defender (Brian)
10,000,000 hours/year testing
→ finds 100,000 bugs/year

Attacker (Paddy)
1,000 hours/year testing
→ finds 1 bug/year

Probability Brian found Paddy's bug: only 10%

Perverse Incentives

Key insight:

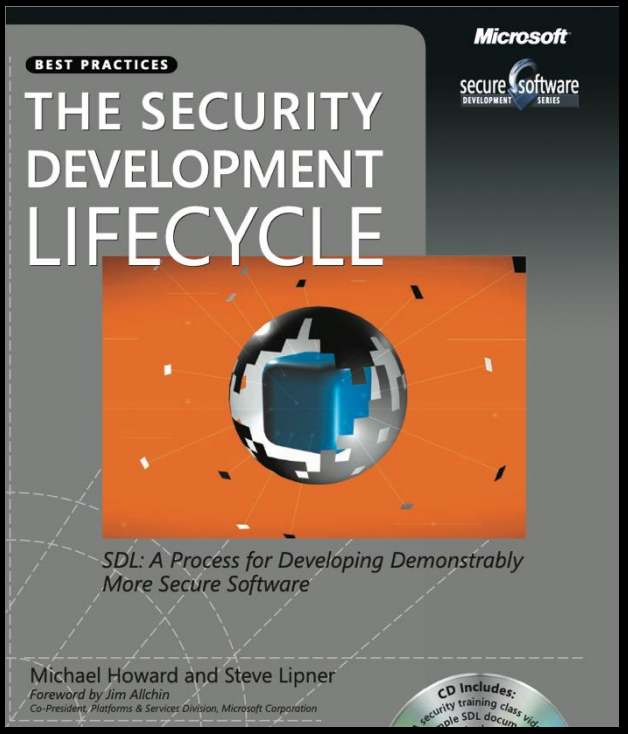
"Information insecurity is at least as much due to perverse incentives" as technical weaknesses.

I.e., economic, not (only) technical

Concept	Definition	Security Example
Network externalities	Value increases with adoption	Platform vendors prioritize developer ecosystem over security; winner-take-all dynamics favor speed over safety
Asymmetric information	One party knows more than another	Buyers cannot assess security quality before purchase ("market for lemons")
Moral hazard	Risk-taking when costs fall on others	Those who make security decisions don't bear the costs of breaches
Adverse selection	Bad products drive out good	Secure products cost more but look identical to insecure ones
Liability dumping	Shifting responsibility to others	Vendors disclaim responsibility via EULAs; costs externalized to users
Tragedy of the commons	Shared resources degraded by individual use	Insecure systems impose costs on others (spam relays, botnets)

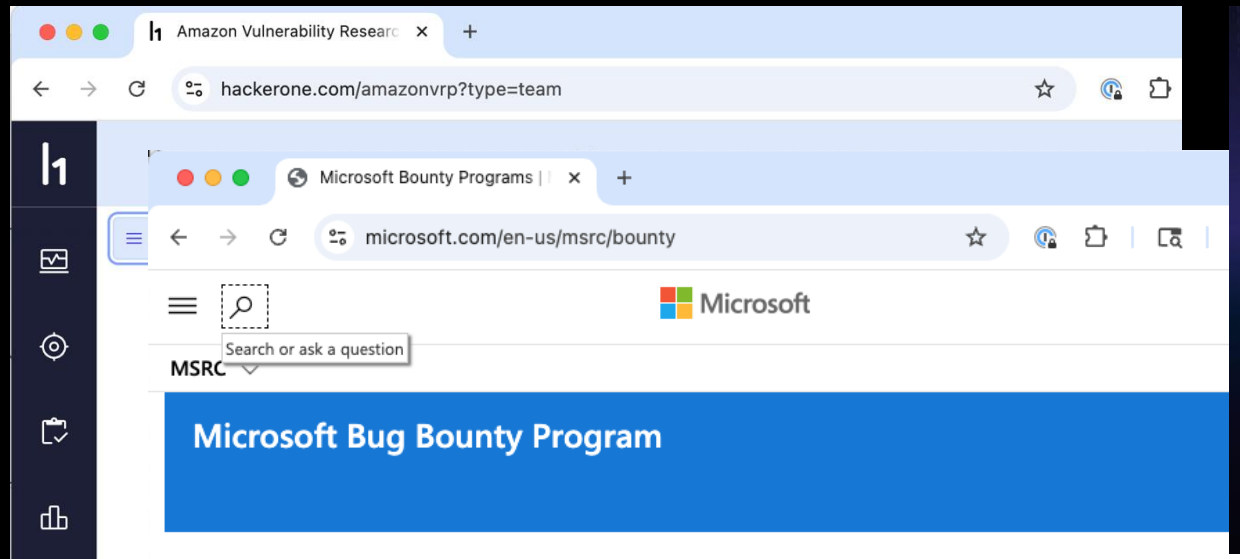
Microsoft Case Study—The Trustworthy Computing Transformation

- 1990s–2000s: Network effects drove Windows dominance
  - "Applications barrier to entry"—security secondary to developer ecosystem
- Code Red, Nimda, SQL Slammer (2001–2003) prompted transformation
  - Bill Gates's "Trustworthy Computing" memo (January 15, 2002)
- Result: **Security Development Lifecycle** (SDL), Windows XP SP2, monthly Patch Tuesday



What has changed for the better?

- Breach disclosure laws (California SB 1386 in 2002, GDPR in 2018) increased transparency
- Bug bounty programs address some information asymmetry
- CISA Secure by Design initiative (2023–2024) pushing vendor accountability

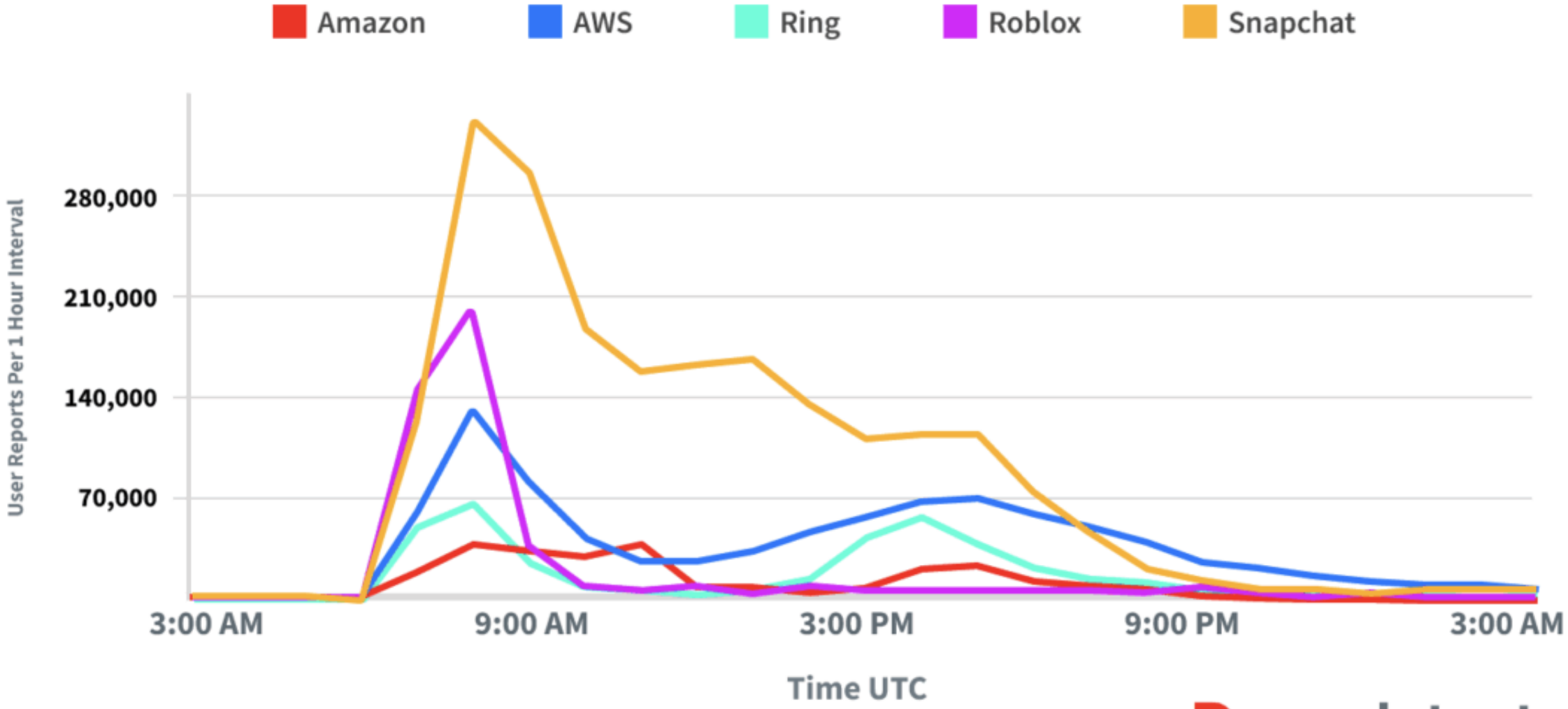


What economic problems have persisted?

Problem	2001 Status	2026 Status
Network externalities	Microsoft dominance	Cloud platform dominance + AI model concentration
Asymmetric information	Can't evaluate security	Still can't evaluate security (now includes AI-generated code)
Moral hazard	Vendors don't bear costs	CrowdStrike: $5.4B damage, ~$0 liability
Liability dumping	EULAs disclaim all	EULAs still disclaim all (though EU CRA may change this)
Tragedy of commons	Spam relays, worms	Botnets, cryptomining, DDoS-for-hire, AI-generated phishing

AWS Global Outage

Downdetector® | October 20-21, 2025



Downdetector
by Ookla

Generative AI as Accelerant and Amplifier

Anderson/Grigg Concept	How GenAI Amplifies It
Attack-defense asymmetry	AI enables personalized attacks at scale; defenders slowed by compliance
Information insufficiency	Neither AI vendor nor developer knows if generated code is secure
Herding	Everyone adopting same AI models creates new monocultures
Network externalities	AI model market is winner-take-all; concentration → systemic risk
Liability dumping	Who's responsible for AI-generated vulnerabilities?

Plus: New attack surfaces

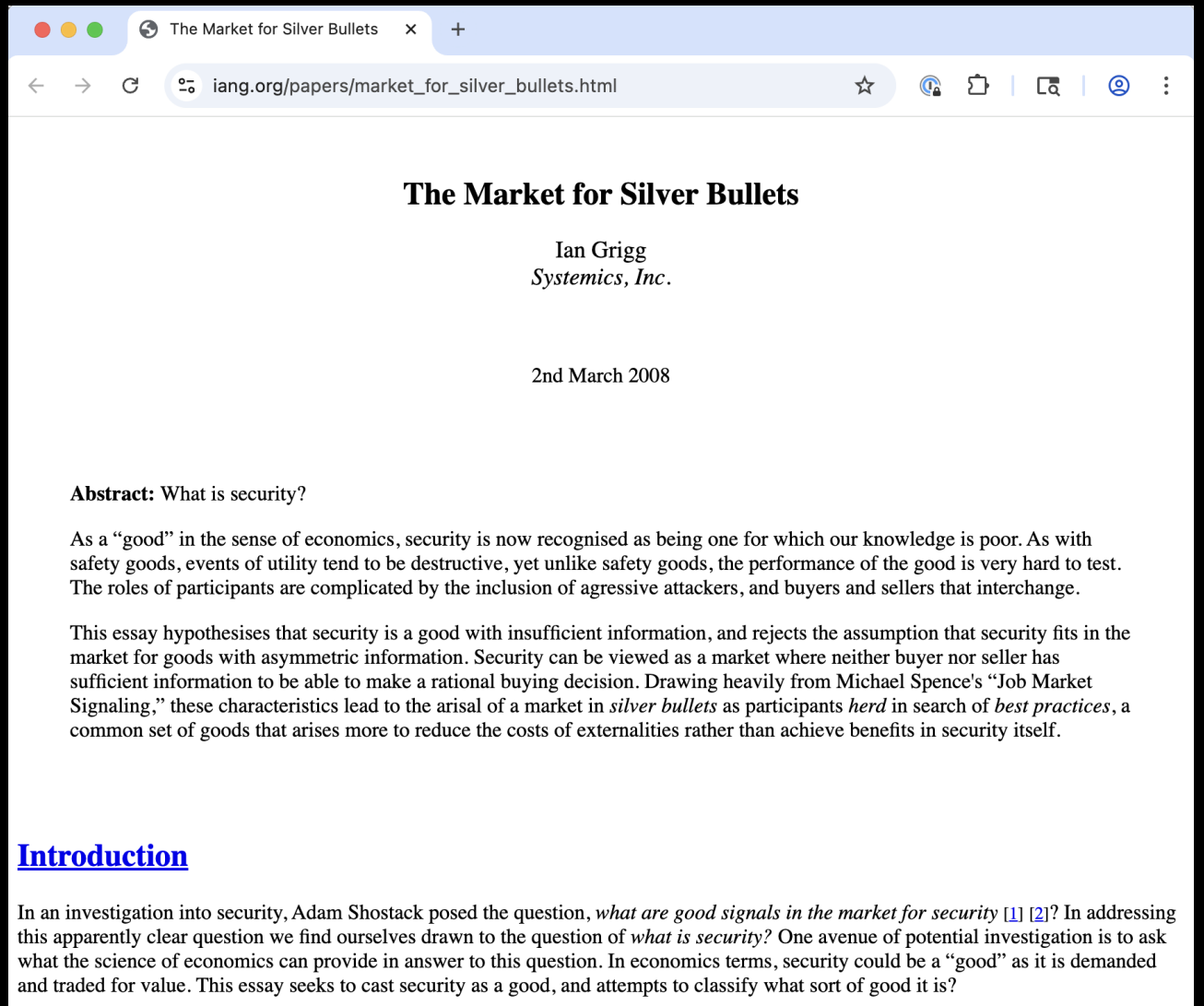
- Prompt injection
- Training data poisoning
- Supply chain attacks via AI coding assistants

What will AI risks stabilize at?

Discussion Questions

- Anderson argues platform vendors prioritize developer convenience over security. **Is this still true?**
- Disclaiming liability: Should software be treated like other engineered products (bridges, aircraft)?
- Does modern evidence still suggest that defenders are inherently disadvantaged?
- Does generative AI tilt the balance toward attackers or defenders, or both equally?
- Will voluntary pledges or liability shifts actually change vendor behavior?

Readings



The Market for Silver Bullets

Ian Grigg
Systemics, Inc.

2nd March 2008

Abstract: What is security?

As a "good" in the sense of economics, security is now recognised as being one for which our knowledge is poor. As with safety goods, events of utility tend to be destructive, yet unlike safety goods, the performance of the good is very hard to test. The roles of participants are complicated by the inclusion of agressive attackers, and buyers and sellers that interchange.

This essay hypothesises that security is a good with insufficient information, and rejects the assumption that security fits in the market for goods with asymmetric information. Security can be viewed as a market where neither buyer nor seller has sufficient information to be able to make a rational buying decision. Drawing heavily from Michael Spence's "Job Market Signaling," these characteristics lead to the arisal of a market in *silver bullets* as participants *herd* in search of *best practices*, a common set of goods that arises more to reduce the costs of externalities rather than achieve benefits in security itself.

Introduction

In an investigation into security, Adam Shostack posed the question, *what are good signals in the market for security* [1] [2]? In addressing this apparently clear question we find ourselves drawn to the question of *what is security?* One avenue of potential investigation is to ask what the science of economics can provide in answer to this question. In economics terms, security could be a "good" as it is demanded and traded for value. This essay seeks to cast security as a good, and attempts to classify what sort of good it is?

- Cryptographer and economist
- Essay published in 2008

Silver bullets, not lemons

Extends Anderson:

- Security is a good
- Security is hard to assess
  - Leads to information insufficiency *on both sides*

The Market for Goods, as described by Information and by Party	**Buyer Knows**	**Buyer Lacks *H1***
Seller Knows	Efficient Goods	Lemons (used cars)
Seller Lacks *H2*	Limes (Insurance)	**Silver Bullets (Security)**

Why Security Can’t Be Tested

“You’re proposing to build a box with a light on top of it. The light is supposed to go off when you carry the box into a room that has a Unicorn in it. How do you show that it works?”

Problem	Explanation
Active attacker	Unlike safety testing, attackers deliberately bypass standardized tests
Burglar alarm paradox	You can test that it beeps, not that it stops burglars
Statistical invalidity	Each attack is unique; past defense doesn’t predict future success
Destructive testing	Real tests are expensive and may cause the harm you’re trying to prevent

Herding and Best Practices

<

Ordinary costs

Extraordinary costs

<u>Herding emerges</u>

- Fingerpointing
- Rational response: Do what everyone else does
- Best practices emerge
- Nash equilibrium

Today: Still herding

- Compliance frameworks as “best practices”
  - SOC 2, ISO 27001, NIST CSF, PCI-DSS
- Security vendor consolidation creating monoculture risk

Today: Still pointing fingers

Post-CrowdStrike outage

- Delta hired David Boies to sue rather than examining their own resilience
- Delta and CrowdStrike sued each other (October 2024)—blame-shifting over root cause analysis

The newest silver bullet: “AI-powered security”

Discussion Questions

- Do you agree that sellers of security products genuinely don't know if their products work against real attackers?
- Herding claim: "Best practices" minimize fingerpointing, not cybersecurity risk; can you identify examples where this is true?
- Do you think it's (still) true that reputational and legal costs of a breach often exceed the direct costs?
- Is AI-assisted coding the ultimate silver bullet scenario, where neither buyer nor seller knows if the output is secure?
- Grigg suggests that transparency, liability reform, and better metrics could break the silver bullet equilibrium. Which interventions are most promising today?

The Role of Regulation

- Both authors suggest market failures require intervention
- What forms of intervention have been tried?
  - Disclosure requirements (SB 1386, GDPR, SEC rules)
  - Standards and frameworks (NIST CSF, ISO 27001)
  - Voluntary commitments (CISA Secure by Design Pledge)
  - Liability reform (EU Cyber Resilience Act)
- What has worked?

Key Takeaways

1. **Security failures are economic, not (just) technical.** Better firewalls won't solve misaligned incentives.
2. **Markets don't self-correct** because neither buyers nor sellers have sufficient information to make rational decisions.
3. **"Best practices" can be a trap**—they emerge from herding dynamics, not effectiveness evidence.
4. **Attack-defense asymmetry is structural**, not incidental. Defenders face thermodynamic disadvantage.
5. **Intervention may be necessary** but must be designed carefully to avoid regulatory capture and new herding equilibria.
6. **The CrowdStrike outage is a case study** in every concept from both papers: network externalities, liability dumping, herding, and the gap between ordinary and extraordinary costs.
7. 🔮 **Generative AI amplifies existing economic problems**—it doesn't create new categories, but intensifies attack-defense asymmetry, information insufficiency, and monoculture risk. The security economics framework from 2001 and 2008 remains essential for understanding AI-era challenges.

Backups

The Data-Driven Approach

This course is about data-driven security management.

How do these economic analyses inform a data-driven approach?

Consider: - What metrics escape the signaling trap? - How do you measure security rather than compliance? - Can economic incentives be aligned through measurement? - How do you measure the security of AI-generated code at scale?

Anderson: Contemporary References

1. **CSIS/McAfee**, "Tilting the Playing Field: How Misaligned Incentives Work Against Cybersecurity" (2017) — Survey of 800 companies showing structural disadvantages for defenders
2. **CrowdStrike outage** — Wikipedia overview; CNN cost analysis (July 2024)
3. **Jen Easterly at Black Hat 2024**: "We don't have a cybersecurity problem, we have a software quality problem" — InsideCyberSecurity coverage
4. **EU Cyber Resilience Act** (entered force December 2024) — First major software liability legislation; official EU page; Wikipedia summary
5. **CSET Georgetown**, "Cybersecurity Risks of AI-Generated Code" (November 2024) — Analysis showing ~50% of AI-generated code contains security-relevant bugs
6. **World Economic Forum**, "Enhancing Cybersecurity Before the Attackers in an AGI World" (October 2025) — On AI amplifying attacker-defender asymmetry

Evaluation and Certification Economics

- Anderson's critique of Common Criteria: "vendor-funded evaluation"
- Evaluator incentive: get paid by the vendor seeking certification
- Modern parallels: SOC 2 audits, penetration testing as compliance checkbox
- **Question**: When does third-party evaluation create real accountability?

The Intelligence Agency Dilemma

Anderson's scenario: A US agency discovers a Windows exploit.

- Report to Microsoft → protect 250M Americans
- Keep quiet → conduct operations against 400M Europeans + 100M Japanese

Credit asymmetry: Operations against foreigners get recognized; defense failures stay hidden

Modern parallel: NSA's Equation Group tools leaked (2016–2017); debate over Vulnerability Equities Process

What Changed Since 2001—Extended Analysis

Positive developments (detail): - California SB 1386 (2002) — first breach notification law; now all 50 states have similar laws - Microsoft's transformation: Trustworthy Computing initiative led to SDL, dramatically reduced Windows vulnerabilities - Bug bounty programs: Mozilla pioneered (2004); now standard at major tech companies - CISA Secure by Design Pledge (May 2024): 194+ signatories; Fortinet year-in-review

Problems that persist (detail): - 83% of organizations report being affected by cybersecurity breaches, yet only 32% report revenue loss (CSIS/McAfee 2017) - Cloud concentration risk: AWS, Azure, GCP control vast majority of enterprise infrastructure - CrowdStrike's 18% endpoint market share created single point of failure (CSA analysis)

GenAI and the Attack-Defense Arms Race

How GenAI amplifies attacker capabilities: - Phishing attacks surged 1,265% since ChatGPT launch (Accenture) - LLM-generated phishing emails achieve 54% click-through rate vs. 12% for human-written (CrowdStrike/Arxiv) - Deepfake incidents up 19% in Q1 2025 alone vs. all of 2024 - $25.6M deepfake fraud at Arup: AI-generated video call impersonated CFO

How GenAI could help defenders: - AI-powered threat detection and anomaly identification - Automated code review and vulnerability scanning - Natural language interfaces for security operations

The adoption asymmetry problem: > “Adversaries are moving faster and experimenting freely with new tools, while defenders are often slowed by bureaucracy, legacy processes and risk aversion.” — WEF 2025

Contemporary References: Silver Bullets

1. **Ross Haleliuk & Mayank Dhiman**, “Cybersecurity is Not a Market for Lemons. It is a Market for Silver Bullets” (Venture in Security, June 2024)
2. **ESPROFILER**, “Herding the Cyber Security Market” (2024)
3. **Cloud Security Alliance**, “What We Can Learn from the 2024 CrowdStrike Outage” (July 2025)
4. **Delta vs. CrowdStrike litigation**: CNN coverage (July 2024); CNBC lawsuit filing (October 2024)
5. **OWASP**, Top 10 for LLM Applications 2025 — Prompt injection ranked #1 risk
6. **Cloud Security Alliance**, “Understanding Security Risks in AI-Generated Code” (July 2025)

The Barings-Visa Paradox

- **Barings Bank**: One breach (Nick Leeson, 1995) → complete collapse
- **Visa**: Millions of fraud events per year → comparatively stable and profitable

Insight: Frequent small failures generate data that enables learning and adaptation; rare catastrophic failures provide no feedback until it's too late.

Application: - Is it better to have many small breaches or gamble on having none? - How does this relate to chaos engineering and resilience testing?

Signaling vs. Substance

Drawing from Spence's job market signaling: - **Education** signals productivity but may not cause it - **Security certifications** signal diligence but may not indicate actual security

Signal	What It Actually Measures
SOC 2 Type II	Auditor found documented controls exist
ISO 27001	Management system is in place
Penetration test	Specific testers didn't find vulnerabilities in limited scope
Bug bounty	Researchers found bugs you're willing to pay for

Question: Which signals, if any, correlate with actual security outcomes?

Breaking the Equilibrium

Grigg's suggestions for escaping silver bullet markets:

1. **Replace signals with metrics**: Develop actual measures of security (ongoing research challenge)
2. **Rebalance costs**: Make direct costs exceed extraordinary costs (liability reform)
3. **Reduce fingerpointing**: Professional norms, information sharing agreements
4. **Sunlight**: Remove secrecy that enables regulatory capture and herding
5. **Incentivize the attacker**: Bug bounties, exploit markets—pay for information about real vulnerabilities

Question: Which of these has been tried? Which has worked?

Modern Silver Bullets and Herding—Extended Analysis

Silver bullets persist: - “AI-powered” security tools with unverifiable claims - “Zero-day protection” promises that can’t be tested until it’s too late - “Next-generation” firewalls, “advanced” threat protection—marketing as signaling

Herding examples: - MITRE ATT&CK evaluations gamed by vendors claiming “100% coverage” (ESPROFILER analysis) - Everyone buying the same Gartner Magic Quadrant leaders - CrowdStrike’s 18% market share created single point of failure

Breach response patterns: - Equifax breach (2017): $700M settlement, stock recovered within 2 years - Target breach (2013): CEO fired, but company thrived - Capital One breach (2019): $80M fine, minimal long-term impact

GenAI Code Generation—A New Silver Bullet Problem?

The scale of AI-assisted development: - 97% of developers have used AI coding tools (GitHub 2024 survey) - “Vibe coding”—trusting AI to handle implementation without careful review

The security reality: - 40–62% of AI-generated code contains security vulnerabilities (CSET Georgetown, November 2024) - Common issues: SQL injection, missing input validation, insecure defaults - Models trained on public repos that contain both secure and insecure patterns

Why this is a silver bullet problem: - Neither the AI vendor nor the developer knows if the generated code is secure - “It compiles and passes tests” ≠ “It’s secure” - Information insufficiency on both sides—classic Grigg scenario

Supply chain amplification: - AI-generated code enters open-source libraries - Vulnerabilities propagate to downstream consumers - Prompt injection in GitHub Actions enables supply chain attacks through AI coding assistants