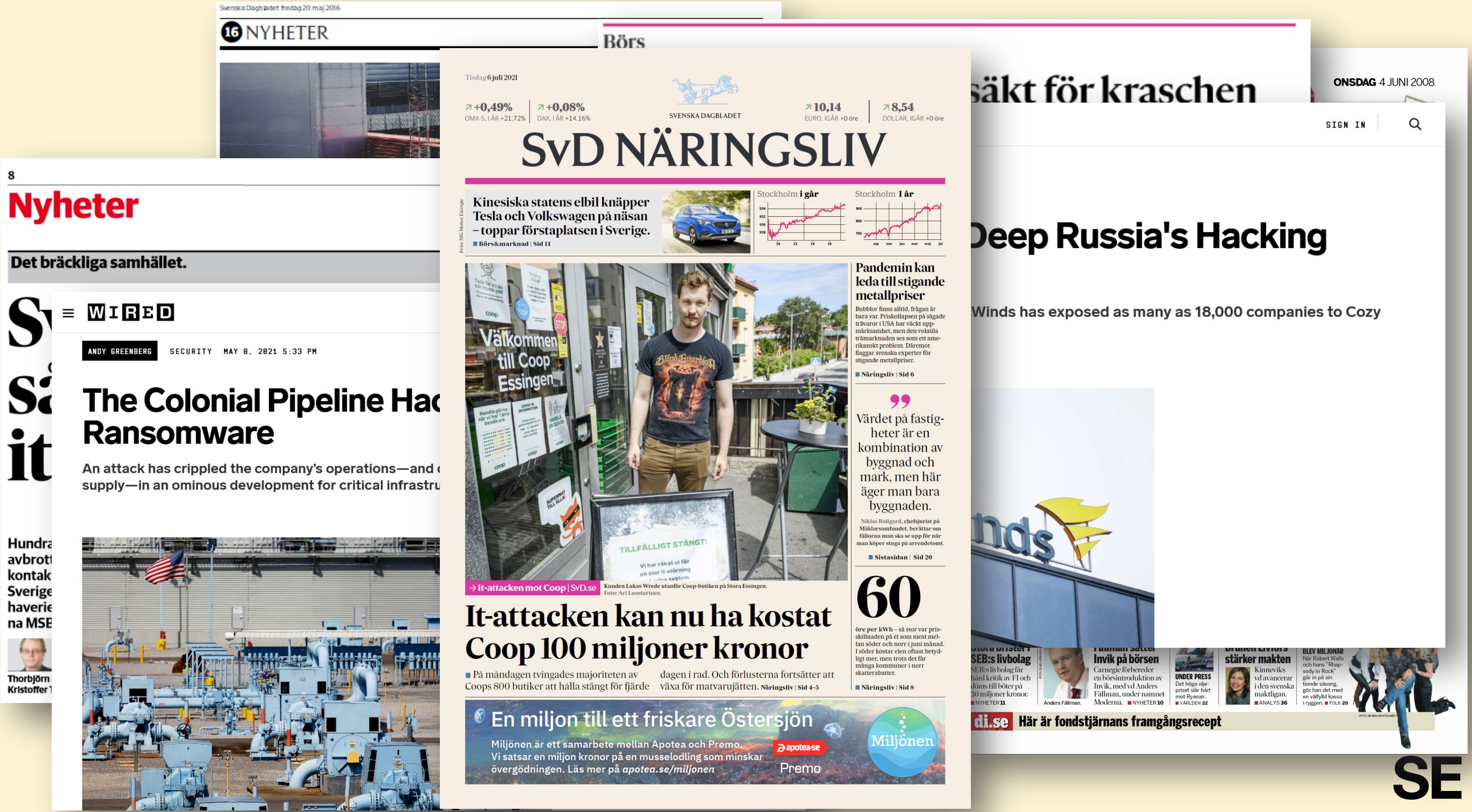
AI, cyber security, and economics

Dr. Ulrik Franke

RISE Research Institutes of Sweden

KTH Center for Cyber Defense and Information Security

We depend on digital services, but are they dependable?



Svenska Dagbladet fredag 20 maj 2016

16 NYHETER

Börs

säkt för kraschen

ONSDAG 4 JUNI 2008

SIGN IN

Deep Russia's Hacking

Winds has exposed as many as 18,000 companies to Cozy

8

Nyheter

Det bräckliga samhället.

Hundra
avbrott
kontak
Sverige
haverie
na MSB

Thorbjörn
Kristoffer

WIRED

ANDY GREENBERG SECURITY MAY 8, 2021 5:33 PM

The Colonial Pipeline Hac
Ransomware

An attack has crippled the company's operations—and c
supply—in an ominous development for critical infrastru

Tisdag 6 juli 2021

SVENSKA DAGBLADET

+0,49% OMX-S, I ÅR +21,72%

+0,08% DAX, I ÅR +14,16%

10,14 EURO, IGÅR +0 öre

8,54 DOLLAR, IGÅR +0 öre

SvD NÄRINGSLIV

Kinesiska statens elbil knäpper Tesla och Volkswagen på näsan – toppar förstaplatsen i Sverige.
Börs&marknad | Sid 11

Stockholm i går

Stockholm 1 år

Välkommen till Coop Essingen

TILLFÄLLIGT STÄNGT!
Vi har råkat ut för en stor it-störning

→ It-attacken mot Coop | SvD.se

Kunden Lukas Wrede utanför Coop-butiken på Stora Essingen.
Foto: Ari Luostarinen

It-attacken kan nu ha kostat Coop 100 miljoner kronor

På måndagen tvingades majoriteten av Coops 800 butiker att hålla stängt för fjärde dagen i rad. Och förlusterna fortsätter att växa för matvarujätten. Näringsliv | Sid 4–5

Pandemin kan leda till stigande metallpriser

Bubblor finns alltid, frågan är bara var. Priskollapsen på sågade trävaror i USA har väckt uppmärksamhet, men den volatila trämarknaden ses som ett amerikanskt problem. Däremot flaggar svenska experter för stigande metallpriser.
Näringsliv | Sid 6

Värdet på fastigheter är en kombination av byggnad och mark, men här äger man bara byggnaden.

Niklas Rollgard, chefsjurist på Mäklarsamfundet, berättar om fällorna man ska se upp för när man köper stuga på arrendetomt.
Sistasidan | Sid 20

60

öre per kWh – så stor var prisskillnaden på el som mest mellan söder och norr i juni månad. I söder kostar elen oftast betydligt mer, men trots det får många kommuner i norr skatterabatter.
Näringsliv | Sid 8

En miljon till ett friskare Östersjön

Miljönen är ett samarbete mellan Apotea och Premo. Vi satsar en miljon kronor på en musselodling som minskar övergödningen. Läs mer på apotea.se/miljonen

apotea.se Premo Miljönen

SEB:s livbolag

Invik på börsen

Anders Fällman.

UNDER PRESS

stärker makten

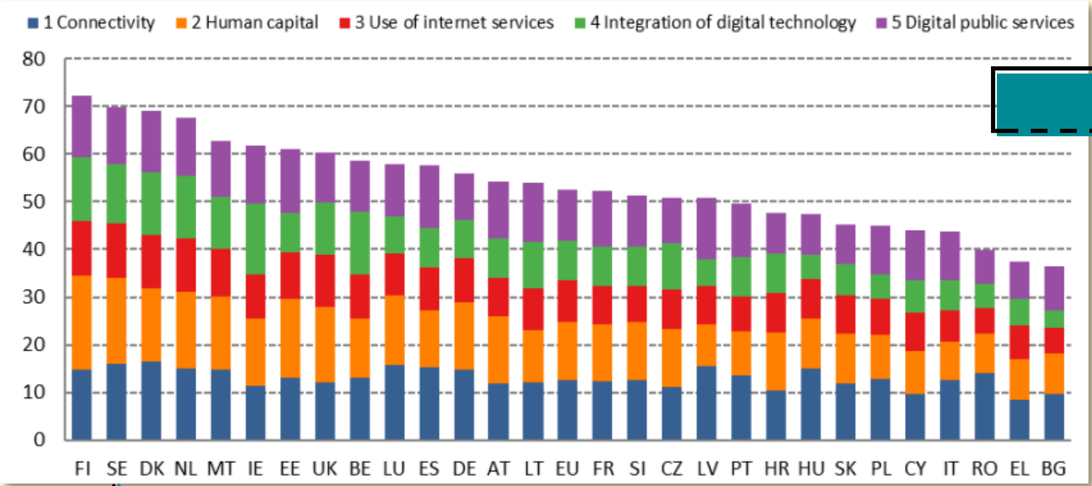
BLEV MILJONÄR

di.se Här är fondstjärnans framgångsrecept

SE

Good news, and bad

Top ranking in the European commission's Digital Economy and Society Index 2020



Only mid-ranked in Europe in the ITU Global Cybersecurity Index 2020

Table 9: GCI results: Europe region

Country Name	Overall Score	Regional Rank
United Kingdom	99.54	1
Estonia	99.48	2
Spain	98.52	3
Lithuania	97.93	4
France	97.6	5
Turkey	97.5	6
Luxembourg	97.41	7
Germany	97.41	7
Portugal	97.32	8
Latvia	97.28	9
Netherlands**	97.05	10
Norway**	96.89	11
Belgium	96.25	12
Italy	96.13	13
Finland	95.78	14
Sweden	94.59	15
Greece	93.98	16
Austria	93.89	17
Poland	93.86	18
Denmark	92.6	19
Croatia	92.53	20
Slovakia	92.36	21
Hungary	91.28	22
Israel**	90.93	23
The Republic of North Macedonia	89.92	24
Serbia	89.8	25
Cyprus	88.82	26
Switzerland**	86.97	27
Ireland	85.86	28
Malta	83.65	29
Georgia	81.07	30
Iceland	79.81	31
Romania	76.29	32
Moldova	75.78	33
Slovenia	74.93	34
Czech Republic	74.37	35
Monaco	72.57	36

The double-edged sword of AI

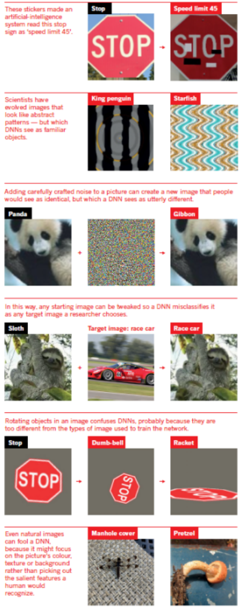
Roll	Medel	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
Data Administration (DTA)	1.00	1.00	1.00	1.00	1.00	1.00
Network Services (NET)	1.14	1.04	1.07	1.48	1.04	1.05
Cyber Operations (OPS)	1.20	1.11	1.14	1.51	1.10	1.12
Cyber Defense Infrastructure Support (INF)	1.26	1.13	1.19	1.74	1.15	1.11
Software Development (DEV)	1.28	1.16	1.22	1.68	1.14	1.20
Cyber Defense Analysis (CDA)	1.29	1.15	1.23	1.78	1.15	1.15
Incident Response (CIR)	1.32	1.19	1.25	1.80	1.18	1.20
Digital Forensics (FOR)	1.37	1.17	1.25	2.09	1.15	1.22
Systems Administration (ADM)	1.41	1.18	1.29	2.21	1.18	1.17
Test and Evaluation (TST)	1.49	1.26	1.38	2.28	1.24	1.30
Systems Analysis (ANA)	1.53	1.28	1.46	2.30	1.24	1.35
Cybersecurity Management (MGT)	1.55	1.27	1.48	2.47	1.26	1.28
Cyber Investigation (INV)	1.55	1.29	1.44	2.46	1.25	1.34
Exploitation Analysis (EXP)	1.56	1.27	1.48	2.48	1.24	1.32
Vulnerability Assessment and Management (VAM)	1.57	1.29	1.45	2.49	1.27	1.32
Technology R&D (TRD)	1.58	1.31	1.46	2.46	1.28	1.37
Systems Development (SYS)	1.63	1.32	1.52	2.62	1.29	1.38
Knowledge Management (KMG)	1.67	1.30	1.52	2.98	1.31	1.27
Systems Architecture (ARC)	1.71	1.34	1.58	2.94	1.32	1.38
Systems Requirements Planning (SRP)	1.73	1.38	1.54	3.00	1.38	1.34
Customer Service and Technical Support (STS)	1.80	1.40	1.76	3.03	1.34	1.46
Risk Management (RSK)	1.82	1.38	1.71	3.25	1.35	1.43
Collection Operations (CLO)	1.83	1.39	1.68	3.31	1.37	1.42
Training, Education, and Awareness (TEA)	1.84	1.38	1.73	3.33	1.35	1.42
Language Analysis (LNG)	1.84	1.38	1.70	3.37	1.35	1.43
Project Management/Acquisition and Program (PMA)	1.86	1.43	1.66	3.33	1.43	1.43
Targets (TGT)	1.88	1.41	1.81	3.33	1.38	1.46
Strategic Planning and Policy (SPP)	2.09	1.48	1.94	4.04	1.46	1.51
Threat Analysis (TWA)	2.10	1.49	2.05	3.95	1.44	1.56
All-Source Analysis (ASA)	2.15	1.51	2.11	4.07	1.46	1.59
Cyber Operational Planning (OPL)	2.41	1.60	2.21	5.01	1.57	1.65
Legal Advice and Advocacy (LGA)	2.41	1.61	2.19	5.02	1.59	1.63
Executive Cyber Leadership (EXL)	2.66	1.68	2.68	5.55	1.64	1.77

Tabell 4: Specialistområden i NICE-ramverket. Siffrorna ska läsas som ett relativt mått på hur lätt det är att automatisera specialistområdet relativt andra specialistområden. Den specialitet som är enklast att automatisera inom respektive modell har index 1,0.

NEWS FEATURE

FOOLING THE AI

Deep neural networks (DNNs) are brilliant at image recognition — but they can be easily hacked.



everything from automated telephone systems to user recommendations on the streaming service Netflix. Yet making alterations to inputs — in the form of tiny changes that are typically imperceptible to humans — can flummox the best neural networks around.

These problems are more concerning than idiosyncratic quirks in a not-quite-perfect technology, says Dan Hendrycks, a PhD student in computer science at the University of California, Berkeley. Like many scientists, he has come to see them as the most striking illustration that DNNs are fundamentally brittle: brilliant at what they do until, taken into unfamiliar territory, they break in unpredictable ways (see 'Fooling the AI').

That could lead to substantial problems. Deep-learning systems are increasingly moving out of the lab into the real world, from piloting self-driving cars to mapping crime and diagnosing disease. But pixels maliciously added to medical scans could fool a DNN into wrongly detecting cancer, one study reported this year[2]. Another suggested that a hacker could use these weaknesses to hijack an online AI-based system so that it runs the invader's own algorithms[3].

In their efforts to work out what's going wrong, researchers have discovered a lot about why DNNs fail. "There are no fixes for the fundamental brittleness of deep neural networks," argues François Chollet, an AI engineer at Google in Mountain View, California. To move beyond the flaws, he and others say, researchers need to augment pattern-matching DNNs with extra abilities: for instance, making AIs that can explore the world for themselves, write their own code and retain memories. These kinds of system will, some experts think, form the story of the coming decade in AI research.

REALITY CHECK

In 2011, Google revealed a system that could recognize cats in YouTube videos, and soon after came a wave of DNN-based classification systems. "Everybody was saying, 'Wow, this is amazing, computers are finally able to understand the world,'" says Jeff Clune at the University of Wyoming in Laramie, who is also a senior research manager at Uber AI Labs in San Francisco, California.

But AI researchers knew that DNNs do not actually understand the world. Loosely modelled on the architecture of the brain, they are software structures made up of large numbers of digital neurons arranged in many layers. Each neuron is connected to others in layers above and below it.

The idea is that features of the raw input coming into the bottom layers — such as pixels in an image — trigger some of those neurons, which then pass on a signal to neurons in the layer above according to simple mathematical rules. Training a DNN network involves exposing it to a massive collection of examples, each time tweaking the way in which the neurons are connected so that, eventually, the top layer gives the desired answer — such as always interpreting a picture of a lion as a lion, even if the DNN hasn't seen that picture before.

A first big reality check came in 2013, when Google researcher Christian Szegedy and his colleagues posted a preprint called 'Intriguing properties of neural networks'[4]. The team showed that it was possible to take an image — of a lion, for example — that a DNN could identify and, by altering a few pixels, convince the machine that it was looking at something different, such as a library. The team called the doctored images 'adversarial examples'.

A year later, Clune and his then-PhD student Anh Nguyen, together with Jason Yosinski at Cornell University in Ithaca, New York, showed that it was possible to make DNNs see things that were not there, such as a penguin in a pattern of wavy lines[5]. "Anybody who has played with machine learning knows these systems make stupid mistakes once in a while," says Yoshua Bengio at the University of Montreal in Canada, who is a pioneer of deep learning. "What was a surprise was the type of mistake," he says. "That was pretty striking. It's a type of mistake we would not have imagined would happen."

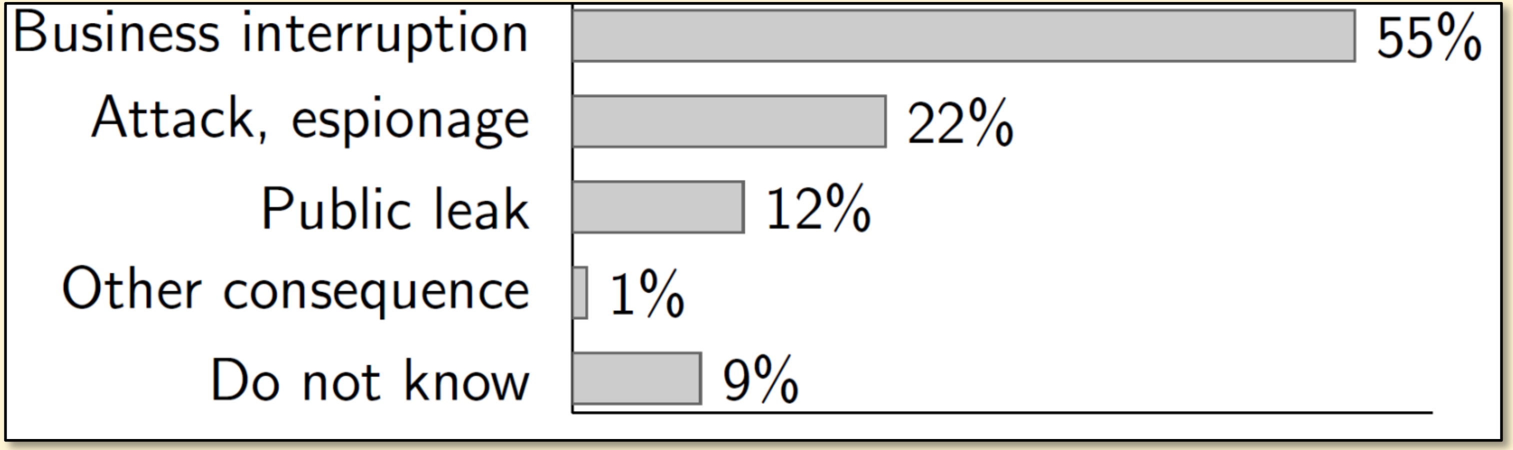
New types of mistake have come thick and fast. Last year, Nguyen, who is now at Auburn University in Alabama, showed that simply rotating objects in an image was sufficient to throw off some of the best image classifiers around[6]. This year, Hendrycks and his colleagues reported

Teodor Sommestad, Joel Brynielsson, Stefan Varga (2019), Möjligheter för automation av roller inom cybersäkerhetsområdet, FOI Memo 6737

Douglas Heaven: Why deep-learning AIs are so easy to fool, *Nature* 574.7777 (2019): 163-166. doi: 10.1038/d41586-019-03013-5

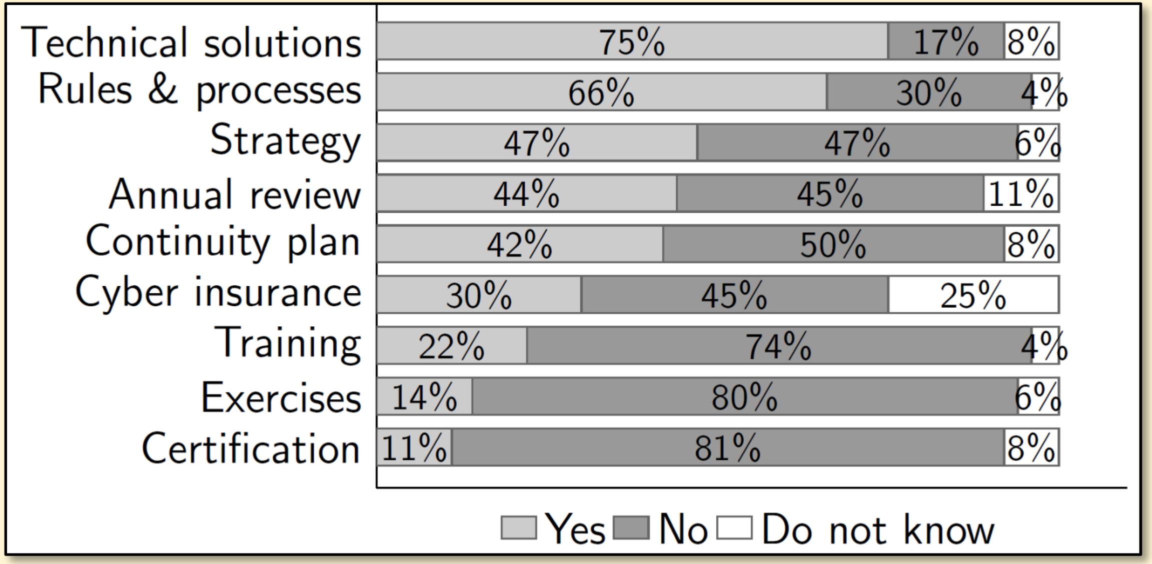
RI. SE

Threat perception in the Swedish manufacturing industry



U. Franke and J. Wernberg, A survey of cyber security in the Swedish manufacturing industry, 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), Dublin, Ireland, 2020, pp. 1-8, doi:

Security measures in the Swedish manufacturing industry



U. Franke and J. Wernberg, A survey of cyber security in the Swedish manufacturing industry, 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), Dublin, Ireland, 2020, pp. 1-8, doi:

Over the past 6 years, people have realized that security failure is caused at least as often by bad incentives as by bad design. Systems are particularly prone to failure when the person guarding them is not the person who suffers when they fail. The growing

Ross Anderson & Tyler Moore: The economics of information security, *Science* 314.5799 (2006): 610-613. doi: 10.1126/science.1130992

The Economics of Information Security

Ross Anderson* and Tyler Moore

The economics of information security has recently become a thriving and fast-moving discipline. As distributed systems are assembled from machines belonging to principals with divergent interests, we find that incentives are becoming as important as technical design in achieving dependability. The new field provides valuable insights not just into "security" topics (such as bugs, spam, phishing, and law enforcement strategy) but into more general areas such as the design of peer-to-peer systems, the optimal balance of effort by programmers and testers, why privacy gets eroded, and the politics of digital rights management.

Over the past 6 years, people have realized that security failure is caused at least as often by bad incentives as by bad design. Systems are particularly prone to failure when the person guarding them is not the person who suffers when they fail. The growing use of security mechanisms to en[illegible]le one system user to exert power over ano[illegible] user, rather than simply to exclude people [illegible] should not be u[illegible]s at all, introdu[illegible] strategic and [illegible] issues. Th[illegible] tools [illegible]epts of game [illegible] and micr[illegible] are becoming [illegible]matics of cryp[illegible]er. [illegible]s and live research [illegible]ics of information se[illegible]ipline is still young, our goal in [illegible] review is to present several promising appli[illegible]ions of economic theories and ideas to practical information security problems rather than to enumerate the many established results. We first consider misaligned incentives in the design and deployment of computer systems. Next, we study the impact of externalities: Network insecurity is somewhat like air pollution or traffic congestion, in that people who connect insecure machines to the Internet do not bear the full consequences of their actions.

The difficulty in measuring information security risks presents another challenge: These risks cannot be managed better until they can be measured better. Insecure software dominates the market for the simple reason that most users cannot distinguish it from secure software; thus, developers are not compensated for costly efforts to strengthen their code. However, markets for vulnerabilities can be used to quantify software security, thereby rewarding good programming practices and punishing bad ones. Insuring against attacks could also provide metrics by building a pool of data for valuing risks. However, local and global correlations exhibited by different attack types largely determine what sort of insurance markets are feasible. Information security mechanisms or failures can create, destroy, or distort other markets; digital rights management (DRM) in online music and commodity software markets provides a topical example.

Economic factors also explain many challenges to personal privacy. Discriminatory pricing—which is economically efficient but socially controversial—is simultaneously made more attractive to merchants and easier to implement because of technological advances. We conclude by discussing a fledgling research effort: examining the security impact of network structure on interactions, reliability, and robustness.

Misaligned Incentives

One of the observations that drove initial interest in information security economics came from banking. In the United States, banks are generally liable for the costs of card fraud; when a customer disputes a transaction, the bank either must show that the customer is trying to cheat or must offer a refund. In the United Kingdom, the banks had a much easier ride: They generally got away with claiming that their automated teller machine (ATM) system was "secure," so a customer who complained must be mistaken or lying. "Lucky bankers," one might think; yet UK banks spent more on security and suffered more fraud. How could this be? It appears to have been what economists call a moral hazard effect: UK bank staff knew that customer complaints would not be taken seriously, so they became lazy and careless. This situation led to an avalanche of fraud (*1*).

In 2000, Varian made a similar key observation about the antivirus software market. People did not spend as much on protecting their computers as they might have. Why not? At that time, a typical virus payload was a service-denial attack against the Web site of a company such as Microsoft. Although a rational consumer might well spend $20 to prevent a virus from trashing his hard disk, he might not do so just to prevent an attack on someone else (*2*).

Legal theorists have long known that liability should be assigned to the party that can best manage the risk. Yet everywhere we look, we see online risks allocated poorly, resulting in privacy failures and protracted regulatory tussles. For instance, medical records systems are bought by hospital directors and insurance companies, whose interests in account management, cost control, and research are not well aligned with the patients' interests in privacy. Incentives can also influence attack and defense strategies. In economic theory, a hidden action problem arises when two parties wish to transact but one party can take unobservable actions that affect the outcome. The classic example comes from insurance, where the insured party may behave recklessly (increasing the likelihood of a claim) because the insurance company cannot observe his or her behavior.

We can use such economic concepts to classify computer security problems (*3*). Routers can quietly drop selected packets or falsify responses to routing requests; nodes can redirect network traffic to eavesdrop on conversations; and players in file-sharing systems can hide whether they have chosen to share with others, so some may "free-ride" rather than help to sustain the system. In such hidden-action attacks, some nodes can hide malicious or antisocial behavior from others. Once the problem is seen in this light, designers can structure interactions to minimize the capacity for hidden action or to make it easy to enforce suitable contracts.

This helps to explain the evolution of peer-to-peer systems over the past 10 years. Early systems proposed by academics, such as Eternity, Freenet, Chord, Pastry, and OceanStore, required users to serve a random selection of files from across the network. These systems were never widely adopted by users. Later systems that succeeded in attracting very many users, like Gnutella and Kazaa, instead allow peer nodes to serve content they have downloaded for their personal use, without burdening them with others' files. The comparison between these architectures originally focused on purely technical aspects: the costs of search, retrieval, communications, and storage. However, it turns out that incentives matter here too.

First, a system structured as an association of clubs reduces the potential for hidden action; club members are more likely to be able to assess correctly which members are contributing. Second, clubs might have quite divergent interests. Although peer-to-peer systems are now thought of as mechanisms for sharing music, early systems were designed for censorship resistance. A system might serve a number of quite different groups—maybe Chinese dissidents, critics of Scientology, or aficionados of sadomasochistic imagery that is legal in California but banned in Tennessee. Early peer-to-peer systems required such users to serve each other's files, so that they ended up protecting each other's free speech. One question to consider is whether such groups might not fight harder to defend their own colleagues, rather than people involved in struggles in which they had no interest and where they might even be disposed to side with the censor.

Danezis and Anderson introduced the Red-Blue model to analyze this phenomenon (*4*). Each node has a preference among resource types—for instance, left-leaning versus right-leaning political

Computer Laboratory, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB3 0FD, UK.

*To whom correspondence should be addressed. E-mail: ross.anderson@cl.cam.ac.uk

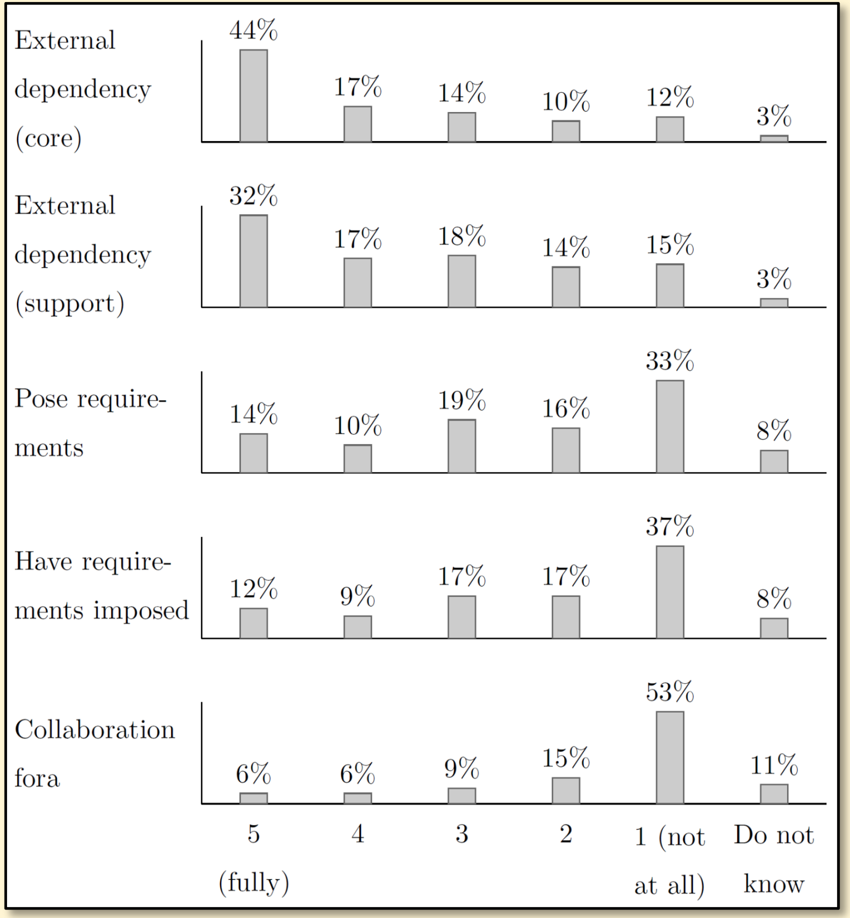
> Over the past 6 years, people have realized that security failure is caused at least as often by bad incentives as by bad design. Systems are particularly prone to failure when the person guarding them is not the person who suffers when they fail. The growing

Cyber security R&D is most often focused on design

If Anderson & Moore are right, as much effort should be spent on incentives as is spent on design!

The promise
and peril
of inter-
connectedness

Ulrik Franke (2020), ”Cybersäkerhet för en uppkopplad ekonomi”, Entreprenörskapsforum



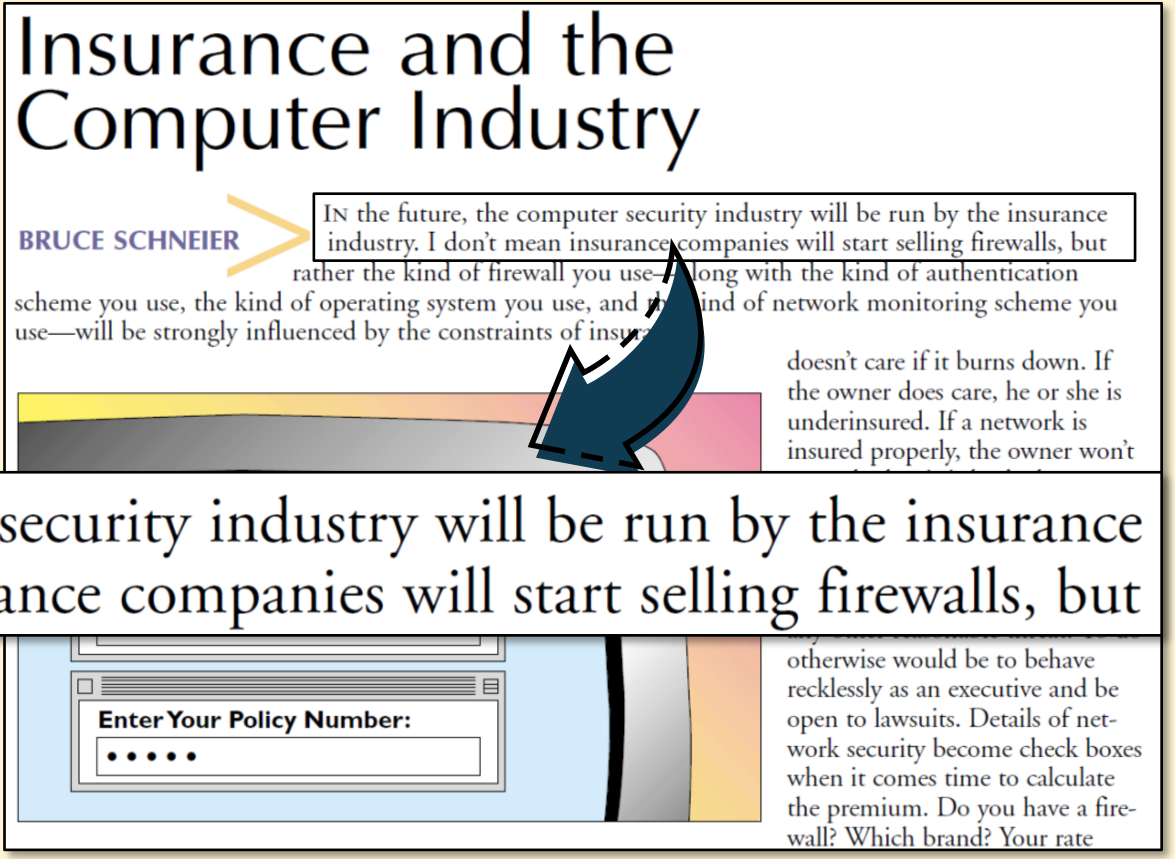
RI. SE

The promise and peril of inter-connectedness (cont’d)

“the companies' overall attitude to sharing vulnerability information is passive but open. In contrast, contemporary cybersecurity guidelines recommend active disclosure and sharing among actors in an ecosystem.”

Olsson, Thomas, et al. "Sharing of vulnerability information among companies–a survey of Swedish companies." 2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE, 2019.

RI.
SE



Insurance and the Computer Industry

BRUCE SCHNEIER

IN the future, the computer security industry will be run by the insurance industry. I don't mean insurance companies will start selling firewalls, but rather the kind of firewall you use—along with the kind of authentication scheme you use, the kind of operating system you use, and the kind of network monitoring scheme you use—will be strongly influenced by the constraints of insur

doesn't care if it burns down. If the owner does care, he or she is underinsured. If a network is insured properly, the owner won't

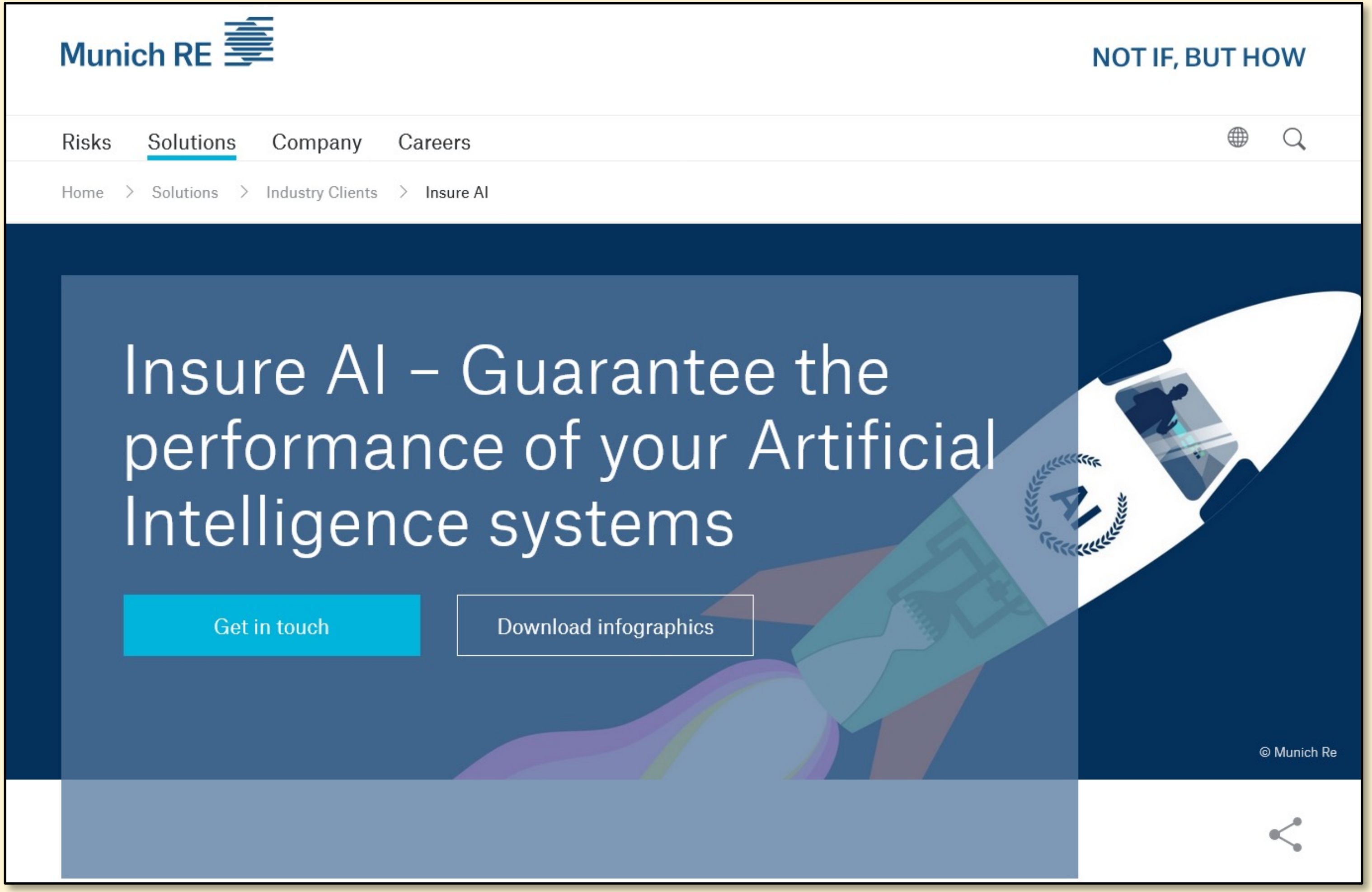
otherwise would be to behave recklessly as an executive and be open to lawsuits. Details of network security become check boxes when it comes time to calculate the premium. Do you have a firewall? Which brand? Your rate

IN the future, the computer security industry will be run by the insurance industry. I don't mean insurance companies will start selling firewalls, but

Bruce Schneier: Insurance and the computer industry, *Communications of the ACM*, 44(3), (2001):114–114. doi: 10.1145/365181.365229
.

Enhancing the Role of Insurance in Cyber Risk Management
OECD

enisa
Cyber Insurance: Recent Advances, Good Practices and Challenges
NOVEMBER 2016
www.enisa.europa.eu
European Union Agency For Network and Information Security


Munich RE
NOT IF, BUT HOW
Risks
Solutions
Company
Careers
Home > Solutions > Industry Clients > Insure AI
Insure AI – Guarantee the performance of your Artificial Intelligence systems
Get in touch
Download infographics
AI
© Munich Re

"I don't think we or anybody else really knows what they're doing when writing cyber… I think anybody that tells you now they know in some actuarial way either what general experience is like in the future, or what the worst case can be, is kidding themselves."

Warren Buffett

Lack of actuarial data

```
0110010010101010101010010110011001101101001010
100?011?1?1?010???0101001??00001?01011???0??10
1?010????10100????1010?1????100?1?0????10100??
??1??1??000???????100????1???0??????100??1????
001??????????10???10??1?11?0001?0?????????????
11010??1?????????????0101????1???100?1????????
????1001?1?1?00?1??????????????1001???????????
???????10010?100???????1010??100?1??100?1?????
?????001?1?01????????????????????0?1??0001????
```

Some knowledge gaps

- **Cyber risks are probably underestimated**
  Attackers stay hidden. Difficult to estimate statistics of rare events (Edwards 2016). Incentives to keep quiet (Bharadwaj et al. 2009).

- **Costs of incidents are great, but we do not know *how* great they are.**
  Surveys are not reliable (Florêncio 2013; Anderson et al. 2013) and incentives are poor (Moore 2010).

- **There is probably an underinvestment problem, but its magnitude is hard to ascertain**
  Negative externalities shift costs to others (Anderson & Moore 2006).

- **We know of many reasonable cyber security measures , but we lack detailed knowledge about their effectiveness**
  Lack of data. There are good ideas (Sonnenreich et al. 2006) but more research is needed.

RI.
SE

?