



Cyberwatch Finland

MAGAZINE 3/2026

Cyber Resilience in 2027

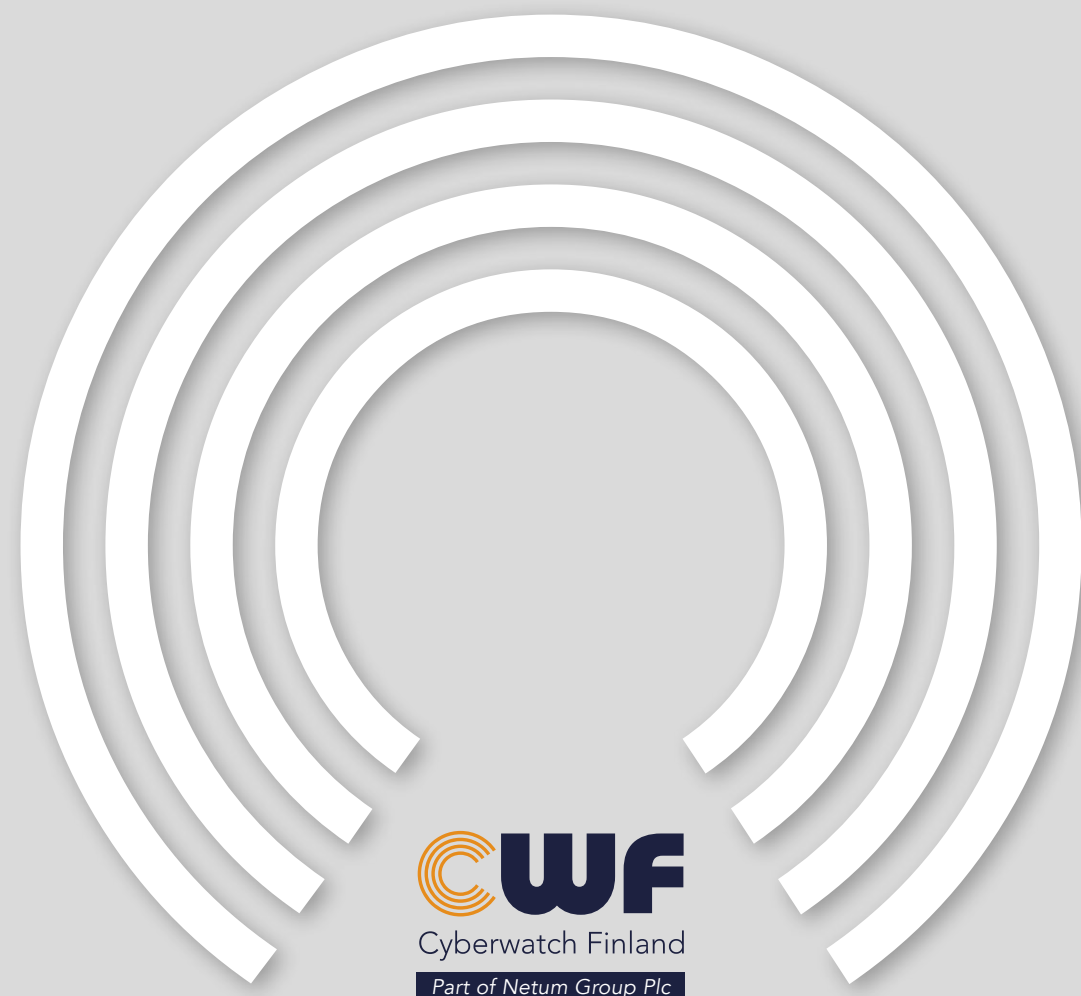


AI GENERATED



Cybersecurity is Built by Small Actions and Management of Large Concepts

➤ Our Aim is to Add
Cyber Capabilities
in the World



CONTENT

4



Editorial
Cyber Resilience
– Solutions for the Future
➤ **AAPO CEDERBERG**

6



Digital Sovereignty
Can Be Lost – Or Won
– Through Everyday
Decisions
➤ **REPE HARMANEN**

10



Cybersecurity Funding
and the Business
Environment Will
Determine Finland's
Security
➤ **RISTO RAJALA
& PETER SUND**

14



Data Poisoning
and RAG Poisoning
➤ **KIMMO K. KASKI**

21



Compliance at the Edge
➤ **SUMMARY OF A CYBERWATCH
FINLAND REPORT
OCTOBER 2026**

27



Cyberwatch Finland
MONTHLY REVIEW
OCTOBER/2026
Cyberwatch Finland became a part of Netum Group Plc 21st of April 2026.

42



Netum & Cyberwatch
CYBER RESILIENCE

Cyberwatch
MAGAZINE

PUBLISHER
Cyberwatch Oy
Kansakoulukatu 1b, Helsinki, Finland

EDITORIAL
Editor-in-Chief Aapo Cederberg
aapo@cyberwatchfinland.fi

LAYOUT
PuulaMedia / Mari Riepponen

ILLUSTRATIONS
AdobeStock, **AI GENERATED**

PRINT
Scanseri Oy, Helsinki

ISSN 2490-0753 (print)
ISSN 2490-0761 (web)

Cyber Resilience – Solutions for the Future

Over recent years, the number of cyberattacks has grown about 30 per cent annually. The use of artificial intelligence in attacks has transformed attack techniques and accelerated both their planning and execution. Vulnerabilities can now be discovered considerably faster, and a tailored cyberattack can be launched within minutes of a new vulnerability being identified. Over the past year, Russian APT groups have made use of AI in 90 per cent of their cyber operations. Nor is AI confined to state-sponsored threat actors; it cuts across the entire spectrum of threat actors.

We must recognise that the challenge is growing and the stakes are getting higher. Future cyber defence must also be able to har-

ness the opportunities offered by AI models, whilst complying with the growing body of regulation governing the use of AI. We must learn to use AI in a responsible way in every domain of cyber defence, threat intelligence, enhancing detection capability, protection and recovery. As a new way of operating, active cyber defence – that is, offensive cyber operations – must form part of a modern cyber defence posture. The cyber equivalent of ‘long-range strike’ is essential if we are to actively degrade attackers’ capacity. Threat intelligence must provide us with timely and anticipatory situational awareness.

In the coming years, Europe will be investing in the development

of conventional military defence, which is both necessary and sensible. Cyber defence has grown in importance as a critical component of military operations, but above all as a guarantor of the functioning of civil society. The impact of wars and crises of varying severity is increasingly directed at society’s critical functions and infrastructure. Cyberattacks, and the amount of them, falling below the threshold of armed conflict represent the fastest-growing area within hybrid warfare operations.

Our society’s development and growth rely on well-functioning digital services. Their continuity must be safeguarded in all circumstances, and particularly in times of crisis. Cyber resilience is becoming ever

more important and going forward we must develop digital services and cybersecurity hand in hand. This is not merely a matter of technological choices, but above all of managing the bigger picture. Do we know our own level of cyber maturity, and are we able to identify our cyber risks? The central role and situational awareness of senior management required under cybersecurity legislation, together with the capacity for risk-based decision-making, will become invaluable. No new ‘tricks’ are needed – what is needed is systematic work to improve the cyber resilience of one’s own organisation.

The use of AI in cyberattacks gives rise to new challenges and threat scenarios. Data is becoming ever

more important, since it is, after all, the basis on which AI operates and ‘thinks’. We have long talked about data manipulation, which occurs continually. A new and insidious phenomenon is now emerging: data poisoning, which may lead, for example, to distorted research findings and errors in operational control processes – with may have devastating consequences to society’s critical functions. This issue features an excellent article by Professor Kimmo Kaski on this new phenomenon, one with which we all must become familiar.

Despite everything, I wish you all a cybersecure end to the year 2026. Let us remember to look after our own organisation’s cyber resilience and to keep improving it.



Digital Sovereignty Can Be Lost – Or Won – Through Everyday Decisions

> REPE HARMANEN,
NETUM GROUP PLC

Finland is at risk of losing its digital sovereignty through decisions that are not regarded as decisions at all. A familiar messaging app, piece of software or cloud platform is adopted in a hurry, and the need for competitive tendering is only noticed once the data has already been migrated. From that point on, it is no longer the technology that is put out to tender, but merely the question of who gets to sell access to a technology that has already been chosen. And, almost without anyone noticing, the organisation's own data has been made subject to the laws of a state outside the EU.

In 2025, the Finnish Competition and Consumer Authority (FCCA) found that VR, Finland's state-owned rail operator, had failed to comply with the Finland National legislation's Act on Public Procurement and Concession Contracts when, in 2023, it procured cloud services and related expert services in six separate packages with a combined value of more than €3 million. VR justified some of the purchases on the grounds that the statutory threshold had not been exceeded, and others on the grounds that there were technical or exclusive-rights reasons why only one supplier could provide the service.¹

At the root of the problem lay VR's IT architecture, which had been built on infrastructure supplied by the technology giant Amazon Web Services. According to the account the company gave to the FCCA, moving to a different cloud technology would have meant rebuilding the entire architecture, which in turn would have required some of the source code for VR's digital applications to be rewritten. The FCCA did not consider this sufficient justification for VR's failure to comply with the law in its procurement.²

In practice, then, VR chose its technology years earlier and had built its IT infrastructure in such a way that it depended on this decision. This now existing dependency simply had never been treated as a decision.

VR is by no means the only contracting entity to have made a choice of this kind, and the same pattern is now emerging in artificial intelligence. Organisations are rapidly adopting commercial language models and AI assistants, building workflows around them and training their staff to use them. Few organisations treat this as a choice of platform, even though that is precisely what it is. And, as in VR's case, such platform choices give rise to digital dependencies that are costly to unwind.

Digital dependency is a problem for the whole of Europe

The market for cloud services, as for AI services, is concentrated in the hands of a few large players. According to the European Commission, the share of the European market held by EU-based cloud providers fell from 29 per cent in 2017 to around 15 per cent in 2022, and has remained at roughly that level ever since. Three technology giants from outside the EU control more than 70 per cent of the market.⁶

In practice, we are all dependent on service providers that are, for the most part, American. In the current geopolitical climate in particular, that is a problem.

The European Commission responded in June 2026 by publishing a strategy package aimed at technological sovereignty, which includes a legislative proposal to support the development of European cloud and AI services: the Cloud and AI Development Act (CADA) ⁵.

CADA sets out four levels for public-sector cloud services, against which their sovereignty could be assessed:

1. The data processed and stored by the service is physically located within the EU.
2. The service provider must be able to demonstrate that it is independent of states outside the EU, and the software supply chain must be transparent.
3. The service provider must be owned within the EU, must comply with EU regulation and must meet further criteria, for instance relating to the nationality of its staff.
4. The service provider's software must be entirely under its own control and transparent, such that countries outside the EU have no means whatsoever of interfering with its operation.

When the package was unveiled, Commission Executive Vice-President **Henna Virkkunen** referred to the risk of an outside party holding a so-called ‘kill switch’ over Europe’s critical services. This fundamental idea lies at the very heart of digital sovereignty. Digital sovereignty does not require every component of our digital infrastructure to be European, but the functioning of critical services must not hinge on the unilateral decision of a party that the EU does not control.⁷

No one should hold a kill switch over European societies.

France has turned the idea into law

In France, the idea is already being put into practice. Since 2021, the state’s use of cloud services has been governed by the ‘Cloud au centre’ doctrine, a policy decision that firmly steers these services towards digital sovereignty. Among other things, the doctrine made the SecNumCloud standard – the set of requirements against which commercial digital services are certified – mandatory wherever a service handles sensitive government data.

The requirements for handling sensitive public data were supplemented by the SREN Act, passed in 2024, and were put into effect by a decree issued in April 2026. In August 2026, the SecNumCloud 3.2 requirements became the official criteria for processing sensitive public data.⁸

In 2025, a French Senate commission of inquiry examined public procurement and its wider effects on the French economy. On the subject of digital procurement, the commission heard evidence from, among others, **Anton Carniaux**, then Director of Legal Affairs at Microsoft France, on public-sector cloud and software services and the dependencies that build up around them. The hearing illustrated very clearly why digital sovereignty is regarded as so important.

Carniaux was asked, under oath, whether he could guarantee that data on French citizens held by Microsoft would never be handed over to a foreign authority without the consent of the French authorities.

His answer: ‘Non, je ne peux pas le garantir.’ I cannot guarantee it.⁹

This is not a question of Microsoft’s morals, but of law. Microsoft is an American company and is ultimately subject to US law. Under the much-discussed CLOUD Act, a US company can be compelled to hand over its users’ data to the government, regardless of the country in which the server holding that data is located.

The same principle applies to other major powers, too. That is why digital security cannot be built on the assumption that a foreign state’s laws will always remain favourable to us.

Finland’s response: a roadmap and a policy statement

In Finland, a ministerial working group of the Government approved a roadmap for digital sovereignty in April 2026. It defines digital sovereignty as the ability to control infrastructure, data and technologies without undue dependence on actors outside the EU. The priorities it identified were mapping critical dependencies, identifying European alternatives and safeguarding the continuity of services.¹⁰

In June 2026, the Government set out its policy on building a national AI platform and computing capacity in order to strengthen digital sovereignty. The policy is particularly timely now that international data-centre operators are arriving in Finland in large numbers, drawn by cheap electricity.

The biggest news in the sector came in September 2026, when Google announced investments of at least €13 billion in Finland over 2027–2028. The investments will

go into digital infrastructure and related projects located in Hamina, Kajaani, Muhos and Vaala.

The investment is significant for Finland, but it does not settle the question of our technological dependencies.³ As Carniaux’s sworn testimony shows, having a server located in Finland does not make a service Finnish in legal terms. US law, for example, extends to data under the control of an American service provider even when that data is stored in Europe.

On top of this, data centres raise the question of electricity. Were all the projects currently planned to go ahead in full, Finland’s electricity consumption would rise by almost 40 per cent on 2025 levels.⁶

Risto Siilasmaa, founder and former CEO of F-Secure, has in his web-article drawn attention to the wider economic perspective. In his view, data-centre investments in Finland are not being examined sufficiently as a whole: how much electricity they tie up, how many permanent jobs they create, and what alternative investments the same electricity might make possible.⁴

Data centres represent investment on an altogether different scale from the technology choices made by companies or contracting authorities, but the same principle holds for both. Decisions are easily made at a rapid pace without paying enough attention to their long-term consequences and effects.

Making digital dependency visible in procurement decisions

The reform of the Finnish Procurement Act that came into force in June 2026 would have offered an opportunity to bring the strengthening of digital sovereignty into everyday procurement practice. The reform’s principal aim was to increase competition: larger contracts must be divided into lots, tendering procedures that attract only

a single bid must in certain circumstances be re-run, and a minimum ownership requirement of ten per cent was introduced for the use of in-house entities. The reform did not, however, tackle digital dependencies.¹¹

Herein lies the difference between Finland and France. Both countries talk about digital sovereignty. France is turning it into binding obligations. In Finland, it remains no more than a roadmap.

The answer is neither isolation nor a ban on American services, but something far more mundane: potential dependencies must be made visible before the procurement decision is taken and contracts are made.

Public-sector’s critical digital procurements ought to include a systematic dependency assessment covering, at the very least, the technical ability to exit, the cost and time involved in doing so, data portability, open interfaces, legal dependency and continuity of service in exceptional circumstances. These factors should be raised as criteria in the evaluation of tenders, alongside price and functionality.

‘We can’t switch because it would cost too much’ is not a law of nature. This is a consequence of the fact that the price of a possible replacement

was never even negotiated. That is why it is all the more important to agree on the right to exit while the supplier still wants the customer – not once the customer wants to be rid of the supplier. The contract must specify data export, formats, documentation, interfaces, migration support, costs and deadlines.

In joint procurement, Hansel – the Finnish state’s central purchasing body – could, for instance, systematically introduce dependency and exit criteria. The next reform of the Procurement Act could go further still and make dependency assessment mandatory, at least for major digital procurements and those critical to the functioning of society.

This does not mean that a Finnish or European supplier should be chosen at any price; rather, it means that technical dependency should be given a price.

Dependence on the goodwill of any single foreign power is a structural weakness, whoever happens to be leading that state at the time. Governments change. Laws change. Sanctions are imposed. Technology is wielded as an instrument of power.

The risk cannot be eliminated by picking the right superpower; it is eliminated by not being entirely at the mercy of any one of them.

REFERENCES/SOURCES

1. Finnish Competition and Consumer Authority, ‘KKV antoi VR:lle huomautuksen laittomasta IT-palveluiden suorahankinnasta’ [FCCA issues VR with a notice for unlawful direct award of IT services], 12 June 2025; decision of 4 March 2025. kkv.fi.
2. Finnish Competition and Consumer Authority, decision R/2024/60, 4 March 2025. kkv.fi/uploads/sites/2/r-2024-60-0024.pdf.
3. Google, ‘Google Deepens Commitment to Finland with Two-Year EUR13 Billion investment in AI Infrastructure’, 9 September 2026. googlecloudpresscorner.com.
4. For Risto Siilasmaa’s assessment of data centres, see Verkkouutiset, ‘Risto Siilasmaa HS:lle: Datakeskukset vievät sähkön, hyöty jää vähäiseksi’ [Risto Siilasmaa to HS: data centres take the electricity, but the benefits are slight], 2026. verkkouutiset.fi.
5. European Commission, Proposal for a Cloud and AI Development Act, COM(2026) 502. eur-lex.europa.eu.
6. European Commission, ‘Strengthening Europe’s tech sovereignty’, 3 June 2026. commission.europa.eu.
7. European Commission press conference, 3 June 2026; remarks by Executive Vice-President Henna Virkkunen on technological sovereignty and the ‘kill switch’ risk to critical services. audiovisual.ec.europa.eu.
8. France: ‘Cloud au centre’ doctrine; SREN Act (2024), Article 31; Décret n° 2026-272, 14 April 2026; confirmation of the SecNumCloud 3.2 requirements in August 2026. legifrance.gouv.fr.
9. French Senate commission of inquiry, hearing of 10 June 2025; Anton Carniaux, Director of Legal Affairs, Microsoft France. senat.fr.
10. Finnish Government, ‘Ministeriöryhmä linjasi digitaalisen suvereniteetin vahvistamisesta – tiekartta ohjaa toimia hallituskauden loppuun’ [Ministerial working group sets out policy on strengthening digital sovereignty – roadmap to guide action until the end of the government’s term], 27 April 2026. valtioneuvosto.fi.
11. Finnish Government, ‘Uusi hankintalaki voimaan – lisää kilpailua julkisiin hankintoihin’ [New Procurement Act enters into force – more competition in public procurement], 2026. valtioneuvosto.fi.

AUTHOR:



REPE HARMANEN
CEO, Netum Group Plc

Repe Harmanen is CEO of Netum Group Plc. Netum is a Finnish IT services company that helps the public sector and businesses build a more effective and secure digital society. Cyberwatch Finland became part of Netum in spring 2026.



➤ RISTO RAJALA
& PETER SUND

Cybersecurity Funding and the Business Environment Will Determine Finland's Security

Political decision-makers, public officials and civil society actors are preparing for the parliamentary elections to be held in spring 2027. The use of public funds once again appears set to become one of the central election themes, as the rise in government debt has not

been brought under control despite the incumbent government's ambitious objectives. At the same time, the deteriorating security situation and commitments related to NATO require higher defence spending, further intensifying the need for adjustment in other sec-

tors of society. Enormous pressure to make savings will inevitably shape the forthcoming government programme negotiations, particularly as all but two parliamentary parties have committed to the jointly agreed parliamentary debt brake. Under this arrangement, they have collec-

tively defined the required level of adjustment while leaving the specific measures to be determined through political decision-making.

The difficult state of public finances also has a direct impact on the development of cybersecurity. The growing attack surface resulting from digitalisation, the accelerating use of artificial intelligence tools in cybercrime and state actors' increasing willingness to exploit information systems for espionage and sabotage place companies and organisations under immense pressure in managing their digital risks—there is no shortage of malicious intent. At the same time, the level of vulnerability remains too high: we estimate that sectors covered by the Cybersecurity Act, such as water supply and energy, each face an investment gap of approximately EUR 100–150 million in reaching a level commensurate with the threats, amounting to as much as EUR 2 billion in total.

Cybersecurity is a core area of security alongside internal security and military defence. The operational reliability of information and communications systems, as well as the confidentiality and integrity of information, are integral to comprehensive security, society's day-to-day functioning, the protection of corporate assets and the economic conditions for business, industrial production and security of supply. Cybersecurity is a fundamental part of economic structures, business processes, production systems and data ownership. Companies and other organisations are on the front line of cybersecurity because they own the systems, identify anomalies, implement practical safeguards and are responsible for recovering from disruptions affecting them. Questions concerning the operational reliability and cybersecurity of ICT solutions form part of every organisation's risk management. Despite the difficulties in public finances—and partly because of them—it is essential that the state provide finan-

cial incentives to companies operating in critical sectors of society and to SMEs that depend on cybersecurity partners so that they can manage cybersecurity risks. Such incentives would address clear market failures, strengthen national security broadly and support the adoption of better technologies. They would also create demand for domestic providers of technology and security solutions. We have therefore proposed that investments to strengthen defence and security be made across administrative branches, with identified needs prioritised jointly by the relevant ministries.

Public opinion supports the objectives we have proposed. According to the Commission's latest Eurobarometer, 96% of Finns believe that the EU and its Member States should strengthen cybersecurity—more than for any other digital policy objective—and 98% consider the protection of information security and privacy online important. In terms of sovereignty, 93% would prioritise investment in digital infrastructure developed and managed in Europe; if they were to change service providers, security and reliability would be the main reason. There is therefore exceptionally strong public support for closing the investment gap in critical sectors, making better use of solutions in public procurement and building the conditions for cybersecurity exports.

In addition to the use of public funds and EU regulation, national legislative processes are under way that, if implemented, will have significant effects on the cybersecurity of society and on the operating conditions of the cybersecurity industry. In domestic policy, over the past year we have focused particularly on procurement and export regulation, the "Act on the Assessment of Information Security", cloud service security and the reform of intelligence legislation. The assessment act and the Military Intelligence Act are discussed in more detail below.

More business opportunities for companies through information system assessments

In 2026, the Government submitted a legislative proposal for an "assessment act" that would amend legislation governing the assessment of information security and preparedness in public authorities' information systems and communications arrangements. The comprehensive reform is welcome: it will strengthen the resilience of public administration, clarify the responsibilities of authorities and improve the availability of assessments. The reform will allow public authorities to make broader use of private experts and companies in support of assessments.

It is important that the assessment and approval of security-critical solutions and the companies that produce them be laid down in law. This will increase transparency, align national procedures with EU and NATO requirements and improve the competitiveness of domestic operators. The change will make the business environment clearer and strengthen confidence in Finnish security technologies. The act will impose an assessment obligation on central government authorities. The obligation to conduct at least a self-assessment and, for certain security classifications, an independent assessment will improve risk management and harmonise interpretations. It will clarify the division of responsibilities and support the secure deployment and maintenance of public authorities' information systems, thereby improving society's overall level of cybersecurity. The proposal assigns the Finnish Defence Forces the role of assessment authority, which is justified in view of the increased need for assessments. During further preparation, however, it should be carefully ensured that this assessment function remains limited and

precisely defined in all respects and covers only the indispensable parts of the Defence Forces' activities and systems.

The reform will also broaden and clarify assessment procedures. To streamline assessments, improve their availability and clarify regulation, the act includes self-assessment by an authority and assessment by a service provider commissioned by an authority. A commissioned service provider could assess information systems that process public information, confidential information and information classified no higher than security class IV. In particular, the option of using a service-provider assessment will improve availability, reduce backlogs and support the maintenance and development of the relevant business activities. Clearer procedures will improve cost-effectiveness, the operational reliability of public authorities and the achievement of assessment objectives within the target timeframe.

It is positive that the procedures for approving the competence of assessment bodies will be made more flexible and clearer. The change will support an adequate supply of assessment services and improve market functioning. For example, it is welcome that TEMPEST companies providing solutions related to compromising emanations will be able to apply for approved status. Verification of the reliability of a company acting as an information security assessment body would be simplified and made more effective by requiring a company security clearance where the body applies for competence to assess the processing of classified information. Using company security clearances in the appropriate situations strengthens reliability. At the same time, it is essential that the Finnish Security and Intelligence Service carefully take account of this change in its operational planning and guidance, so that the company security clearance procedure does not become a de facto bottleneck.

The impact of amendments to the Military Intelligence Act on companies as an example

The Government's proposal to amend the Military Intelligence Act was considered by Parliament in autumn 2026. The proposal aims to make the intelligence activities of the Finnish Defence Forces more effective and to safeguard national security through modern powers. Its objective is justified, as the powers of military intelligence must reflect the security environment and technological development. However, the proposal affects companies exceptionally directly, including data centres, communications intermediaries, hardware and software manufacturers, and companies whose systems could also be used as intelligence tools.

During the further preparation of the proposal, several shortcomings previously raised by the business community have been addressed.

The explanatory memorandum now states that companies will not be required to dismantle end-to-end encryption or install backdoors, and that encryption is essential to Finland's comprehensive security. It is also important that the search criteria for communications intelligence be specified more precisely and that the duty to assist be limited to data centre service providers covered by the Cybersecurity Act. The clarifications made to the authorisation procedure for intelligence directed at the information system of a state actor within Finland are also necessary.

The main unresolved issue in the reform concerns the position of companies as targets of intelligence measures. If a company detects software installed by military intelligence or an unfamiliar device in its system, this constitutes an information security incident that may trigger a reporting obligation under the Cybersecurity Act. The company could therefore be required to investigate, report and remediate a measure taken by the state itself, while bearing the costs and reputational risk. It must be clear that such an incident cannot result in sanctions against the company or its management. Communications intelligence concerning a non-state actor should be disclosed to a person legally residing in Finland and to a company operating lawfully in Finland, as the Constitutional Law Committee has found that a notification obligation could better safeguard legal protection.

The proposal also addresses measures targeting information systems located outside Finland. We have taken the view that such measures should be decided at the level of the state's political leadership and that their use should be limited to countering a serious and immediate threat. This is because possible countermeasures would likely affect critical infrastructure owned by companies and their business opportunities. The participation of a company employee in assisting intelligence activities should also require

the employer's knowledge and consent. The combined effects of amendments to intelligence legislation must also be assessed as a whole, because the reform is being considered alongside a forthcoming proposal concerning residential premises that would permit the use of intelligence powers in homes—which are also remote-work offices. We have also required that the limits and oversight of the use of cover in open-source intelligence gathering be laid down in law. Information obtained from a private entity must not enable the requirements governing covert intelligence-gathering methods to be circumvented. Precisely defined and supervised regulation strengthens confidence in Finland as a safe and predictable operating environment. This is important not only for Finland as a location for cybersecurity companies, but also for its position as a hub for data centres and telecommunications infrastructure.

Towards the parliamentary elections

A successful, growing and internationalising cybersecurity industry is in many ways indispensable to our country's security and prosperity and to maintaining the functioning of our society. It is essential that the next government, whatever its composition, be able to find the resources needed to maintain and develop cybersecurity while safeguarding the operating conditions of business. At the same time, legislative and administrative processes must not grant public authorities powers that would weaken the business opportunities of companies operating in the market or jeopardise the fundamental-rights protection of property, privacy and confidential communications. Guaranteeing and strengthening the confidentiality of communications and well-functioning markets are vital to the domestic cybersecurity industry and, more broadly, to our country's digital economy.

Finnish Information Security Cluster's advocacy work will continue actively. We will continue to highlight issues essential to the cybersecurity industry and to society's cybersecurity, both in public debate and in meetings with political decision-makers, public officials and other stakeholders.

AUTHORS:



PETER SUND

CEO

Finnish Information Security Cluster (FISC)
Technology Industries of Finland



RISTO RAJALA

Advisor

Finnish Information Security Cluster (FISC)
Technology Industries of Finland





Data Poisoning and RAG Poisoning

The article is based on a memorandum and a presentation of it, presented at King's College London in the summer of 2026.

Data poisoning is an **integrity** attack on shared data: rather than stealing information, the adversary corrupts it so that any model trained on, or any system retrieving from, that data inherits the attacker's intent. It is uniquely relevant to open and collaborative research, whose value rests on trusting data one did not create. The contamination needed is small and its realisation is cheap: replacing roughly **0.001% of training**

tokens was enough to corrupt a medical Large Language Model's (LLM) answer [2], the number of poisoned documents needed is **near-constant** – **about 250** – **regardless of model size** [24], and in Retrieval-Augmented Generation (RAG) as few as five injected documents reached a 90% attack success rate against a million-document store [13]. The third update (October 2026) adds compositional and 'knowledge-update' poisoning [39–44], poisoning beyond question answering and of agent memory [45–49, 54], the first provably robust aggregation defence for RAG [53], and the EU Digital Omnibus on AI [59].

1. Poisoning as an Integrity Threat to Shared Data

Unlike confidentiality breaches, poisoning targets the trustworthiness of data and the models that consume it. It is **stealthy** (overall metrics can look healthy while a narrow behaviour is corrupted), **cheap** (a tiny poisoned fraction suffices) and **durable** (it persists in weights or in a knowledge base until detected).

The open-science channels that make data valuable are also the attack surface: crowdsourced datasets, public repositories and the supply chain, web-scraped pretraining

corpora, federated learning and even peer review. Appending as few as 13 words to one popular user-generated page got attacker text cited in 38–51% of deep-research agents' reports [46].

BadRAG showed that ~0.04% of a corpus could be retrieved with 98.2% success and raise a GPT-4 RAG refusal rate from 0.01% to 74.6% [15]. The ~250-document threshold held across models from 600M to 13B parameters – only 0.00016% of tokens for the largest – overturning the assumption that larger models require proportionally more poison [24].

2. Threat Landscape: Classic Data Poisoning

Attack goals fall into three groups. **Availability** (indiscriminate) attacks degrade accuracy broadly [1]; **integrity** (targeted) attacks corrupt behaviour on a narrow slice while leaving aggregate metrics intact; and **backdoors** embed a rare trigger, so the model behaves normally until the trigger appears, keeping standard test sets clean [3].

Methods range from **label flipping** to **clean-label** and **optimisation-based** poisons; the asymmetry between tiny poisoned fractions and

large behavioural effects is the defining risk [2, 24]. Results from 2026 raise the bar: the choice of poison set alone moved backdoor success on an 8B LLM model from 3% to 80% [50], Phantom Transfer survives eleven data-level defences [51], and backdoors are becoming programmable [57] – so random-sample evaluations understate the worst case.

Entry points include data collection, curation, packaging and distribution, continual updating, the ingestion toolchain [28] and decentralised fine-tuning, where one node that poisons its own adapter can backdoor its peers [55].

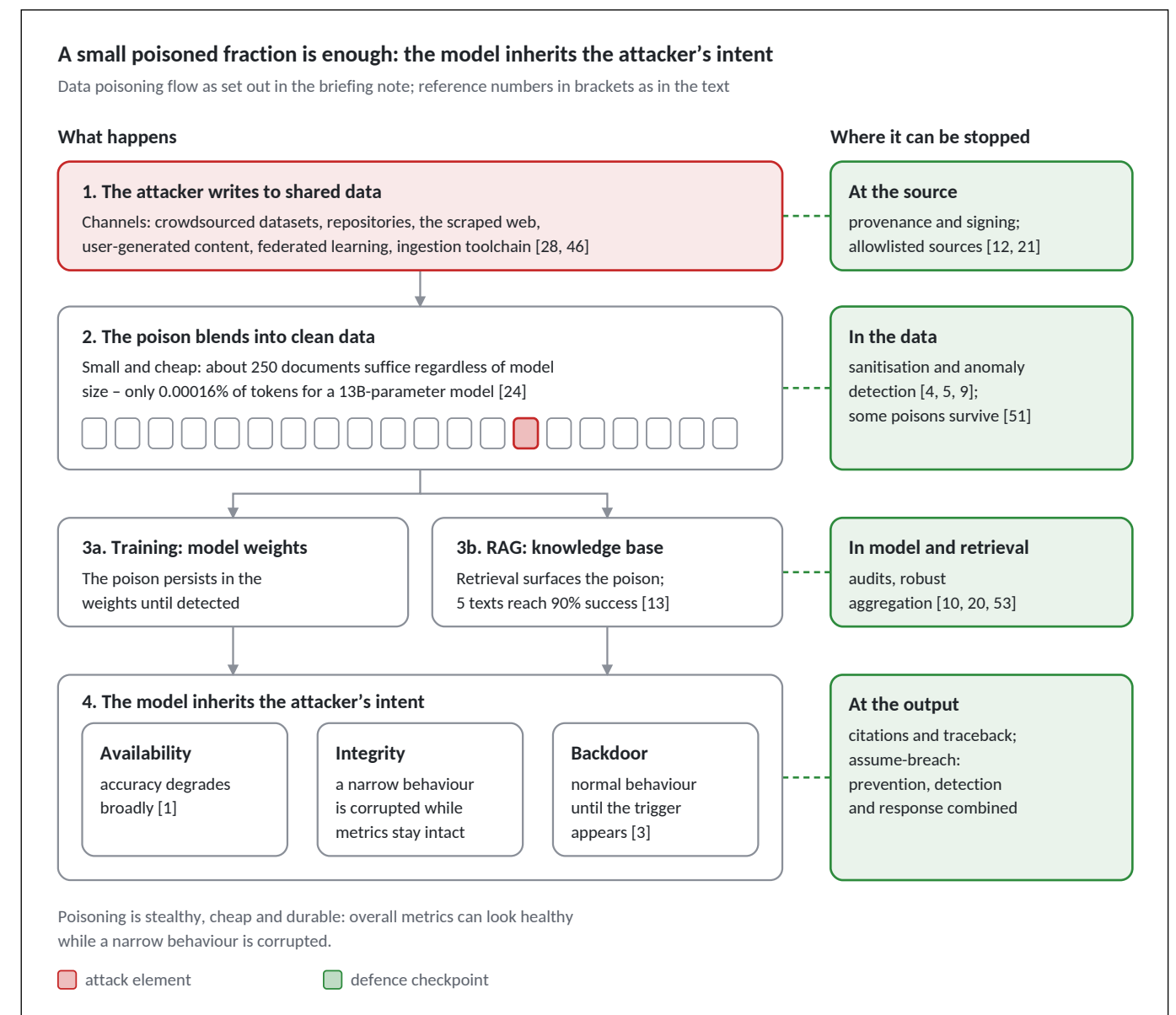


Figure 1. Data poisoning flow from entry to impact, with defence-in-depth checkpoints.

3. RAG Poisoning:
The Knowledge-Base
Attack Surface

The RAG knowledge base, used by most LLM models, is a new and prac-

tical attack surface: an adversary with direct or indirect write access can inject documents the retriever will surface and the model will trust [13]. A poisoned document must be **retrievable** and **steering** – carry-

ing the false answer or instructions the model obeys as **indirect prompt injection**. Targets now extend to agent memory, code generation [45], security tooling [54] and deep-research agents [46, 47].

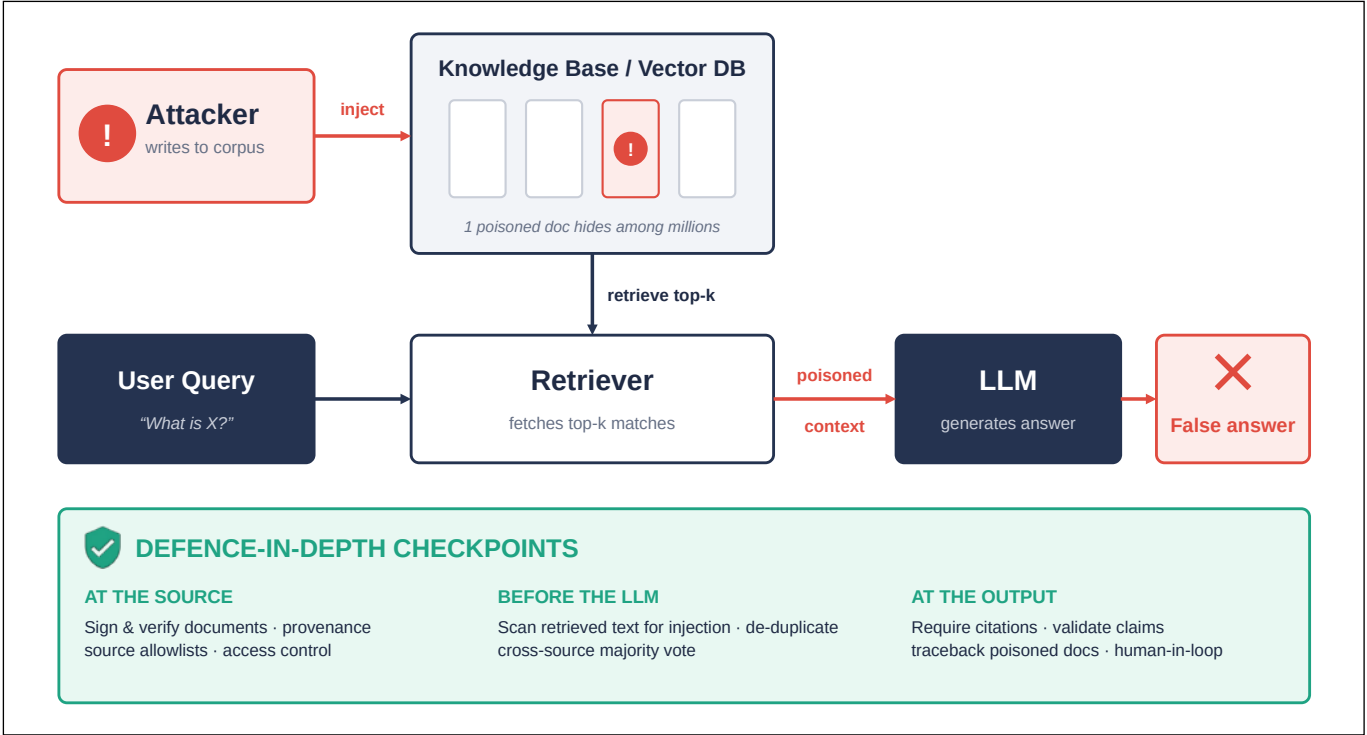


Figure 2. RAG poisoning flow and defence-in-depth checkpoints.

4. RAG Attack Taxonomy

Attacks can be classified along five orthogonal dimensions: **pipeline phase** (retrieval or generation [15]), **access** (black-box, white-box or, strongest, retriever fine-tuning), **query dependence**, **crafting method** (heuristic, optimisation-based [14] or compositional [39, 40]) and **goal** (targeted answer, denial of service, steering, jailbreak bypass, data exfiltration or insecure code [45]).

5. Threat-vs-Control
Matrix

The matrix pairs each threat with detection, prevention and response controls. No single control is sufficient; the intent is layered, defence-in-depth coverage. The original’s 18 threats are grouped here into seven.

6. Detection Methods

For training data: spectral signatures [4], activation clustering [5], influence-based attribution [9], trig-

ger reverse-engineering [6], STRIP [7], fine-pruning [8] and canary examples that should be learned as expected. **For RAG:** retrieval and embedding anomalies, scanning retrieved chunks for prompt injection or hidden text, cross-source consistency checks and knowledge-base traceback to the responsible documents. Poisoned answers can have lower perplexity than benign ones, but attention ‘collapses’ onto the poisoned documents, which D-SCAN detects even when the answer does not change [52]; soft trust demotion outperforms hard filtering [54].

ATTACK	PHASE; ACCESS	PRIMARY GOAL
PoisonedRAG [13], CorruptRAG [31], Confundo [30]	Retrieval + generation; black-box	Targeted, pipeline-robust answer
BadRAG / TrojRAG [15]	Retrieval-phase backdoor; white-box	Denial of service / semantic steering
HijackRAG [17]	Generation; black-box	Answer hijacking
Corpus poisoning [14], MIRAGE [29], POISONCRAFT [19]	Retrieval; query-agnostic	Retrieval pollution / misinformation
Poisoned-MRAG [18] / One Pic [28]	Multimodal; optimised image	Targeted answer / disinformation
GRAGPoison [28], KG-RAG / KEPo / LogicPoison [32–34]	GraphRAG; graph-structure-aware	Reasoning-path hijacking
DisarmRAG [35]	Retriever editing; white-box	Self-correction suppression
InceptionRAG [39], micro-collaborative poisoning [40]	Compositional; several passages / sources	Induced false conclusion
ToxicRAG / PURPOSE / DenialRAG / RAG-NAROK [41–44]	Conflict resolution; single document	Evading conflict-resolution defences
CodePoisonRAG [45]	Code knowledge base; ≤1 artefact	Insecure code (CWE) injection
UGC / web-evidence poisoning [46, 47]	Live web search	Report and citation manipulation
AgentPoison [16], GhostWriter [48]	Agent memory; trigger / hidden payload	Persistent malicious action

Table 1. Representative RAG poisoning attacks by taxonomy dimension (condensed).

THREAT	DETECTION	PREVENTION	RESPONSE
Training-data poisoning (availability, integrity, backdoor)	Subgroup monitoring; influence functions; spectral signatures; STRIP	Provenance + signing; curation; robust training	Retraining; fine-pruning; unlearning
Federated learning and decentralised fine-tuning	Update-distribution anomalies; adapter screening	Robust aggregation (Krum, trimmed mean); Chorus	Node exclusion; re-aggregation
Supply chain and ingestion	Hash and signature mismatches; parser-level inspection	Signed artefacts; data/model BOM; trusted loaders	Quarantine; rollback to verified version
RAG knowledge base and prompt injection	Retrieval and embedding anomalies; instruction-pattern scanning; traceback	Allowlisted sources; retrieved text treated as untrusted	Removal; re-indexing; multi-source re-answering
Compositional, learned and conflict-evading poisoning	Attention anomalies (D-SCAN, AV Filter); cross-passage consistency	Document isolation (HODOR); PRA-RAG; authority-based reranking	Whole-chain traceback and removal; independent re-verification
GraphRAG and code knowledge bases	Topology anomalies; CWE and static analysis	Triple provenance; signed code knowledge bases	Triple pruning; audit of generated code
Agents: tools, memory, web search, review	Memory integrity checks; cited-source monitoring	Least privilege; signed memory; UGC down-weighting	Memory rollback; report regeneration; human review

Table 2. Threat-vs-control matrix (condensed).

7. Defence and Mitigation (Defence-in-Depth)

- **Provenance & integrity:** track lineage; hash and sign data and model artefacts; version control to diff and roll back. Provenance is what makes attribution and recovery possible [12, 21].
- **Source & supply-chain hygiene:** prefer authenticated, reputable sources; avoid indiscriminate scraping; vet third-party datasets and models via model cards and lineage.
- **Sanitisation & robust training:** robust aggregation – Krum [10], trimmed mean / median [11]. In decentralised fine-tuning, Chorus cut attack success from 48–63% to ≤2.2% [55], and unlearning-based decontamination removed most agent backdoors [56]. Because some poisons survive every data-level filter tested [51], pair sanitisation with post-training model audits.
- **RAG-specific controls:** allowlisted sources; retrieved text treated as untrusted (never execute embedded instructions); post-retrieval filtration (FilterRAG) [25]; isolate-then-aggregate generation (RobustRAG) [20]; sparse document attention (SDAG) [27]; attention-aware detection (AV Filter) [36]; information-flow control (Cordon-MAS) [37]; signed agent memory (SMSR) [38]; provably robust aggregation (PRA-RAG) [53]; document isolation (HODOR) [39]; memory retrieval screens (AM-Sentry) [48]; and, for deep-research agents, limiting reliance on single user-generated pages – dropping them entirely blocks the attack but costs report quality [46].
- **Access control & audit:** restrict who can write to datasets and knowledge bases; keep tamper-evident audit logs.

8. Governance and Standards

Anchor the programme in recognised frameworks: NIST AI 100-2

[21], the OWASP Top 10 for LLM Applications (LLM04:2025) [22] and, for agentic systems, ASI06 Memory & Context Poisoning [58], and the EU AI Act, which makes robustness against poisoning a compliance obligation for high-risk systems (Articles 10 and 15) [23]. The Digital Omnibus on AI (Regulation (EU) 2026/1744, in force since 27 July 2026) defers these requirements to 2 December 2027 for Annex III systems and 2 August 2028 for AI embedded in regulated products [59]. The obligations are deferred, not removed, so the interval is best used to build provenance and robustness evidence.

9. Recent Developments (2025–2026)

2025. The most consequential result is that poisoning requires a **near-constant number of documents** regardless of LLM model size [24]. RAG attacks broadened to query-agnostic, graph, multimodal and ingestion-toolchain poisoning [18, 19, 28, 29], and defences concentrated on the retrieve–generate boundary [20, 25, 27] – yet common mitigations only partially blunt the strongest attacks [13, 18].

January–June 2026. Confundo showed that learning-based ‘pipeline-robust’ poisons outperform purpose-built attacks by up to ~1.7× even under defences [30, 31]; structured RAG became a first-class target [32–34]; and DisarmRAG poisons the retriever itself, suppressing the LLM’s self-correction with >90% success [35].

July–October 2026. Poisoning became **compositional** (InceptionRAG >80% success [39, 40]); single-document attacks evade conflict resolution by posing as knowledge updates (ToxicRAG 61–91% [41–44]); CodePoisonRAG reaches 80–93% success at a 0.7% poisoning ratio [45]; and malicious agent memory persisted in 84% of cases [48, 49].

Defences 2026. Attention-aware methods recover up to ~20% accu-

racy, yet adaptive attacks still reach ~35% success [36]; PRA-RAG reduces attack success to as low as 1% while keeping 71% accuracy [53]. The caveat stands: each new attack family was shown to evade at least some existing defences [39, 41–45], so gains remain defence-specific rather than general.

10. Residual Risk and the Openness Tension

Defences are imperfect: standard mitigations only partially help [13, 18], detector performance is highly sensitive to threat model and configuration, and the near-constant-count finding means scale offers no inherent protection [24]. The realistic posture is **assume-breach**: combine prevention, detection and response rather than relying on any one line. Measured ‘lab’ success rates under- or over-state real exposure depending on the pipeline [30, 31], and adaptive attackers partially conceal attention-level signals [35, 36] – an arms race, not a solved problem. Certified defences now exist for narrow settings but not for general robustness.

The July–October 2026 results add two lessons. Poisons that are individually benign but collectively harmful defeat per-document inspection, so controls must reason over the retrieved set as a whole [39, 40]. And agents that browse the open web or keep long-term memory inherit the integrity of every page and conversation they ingest, which extends assume-breach from the knowledge base to the agent’s whole information diet [46–49].

For academia this creates a real tension with open-science and FAIR norms: the same openness that makes data valuable makes it poisonable. The mitigations that scale best are the ones that preserve openness – signed and versioned datasets, transparent provenance, reproducibility (which doubles as detection), and clear curation and review responsibilities – rather than closing data off.

REFERENCES/SOURCES

[1] Biggio, B., Nelson, B., & Laskov, P. (2012). Poisoning Attacks against Support Vector Machines. ICML.

[2] Alber, D. A., et al. (2025). Medical large language models are vulnerable to data-poisoning attacks. Nature Medicine, 31(2), 618–626. doi:10.1038/s41591-024-03445-1.

[3] Gu, T., Dolan-Gavitt, B., & Garg, S. (2017). BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain. arXiv:1708.06733.

[4] Tran, B., Li, J., & Madry, A. (2018). Spectral Signatures in Backdoor Attacks. NeurIPS.

[5] Chen, B., et al. (2018). Detecting Backdoor Attacks on Deep Neural Networks by Activation Clustering. arXiv:1811.03728.

[6] Wang, B., et al. (2019). Neural Cleanse: Identifying and Mitigating Backdoor Attacks in Neural Networks. IEEE S&P.

[7] Gao, Y., et al. (2019). STRIP: A Defence Against Trojan Attacks on Deep Neural Networks. ACSAC.

[8] Liu, K., Dolan-Gavitt, B., & Garg, S. (2018). Fine-Pruning: Defending Against Backdooring Attacks on Deep Neural Networks. RAID.

[9] Koh, P. W., & Liang, P. (2017). Understanding Black-box Predictions via Influence Functions. ICML.

[10] Blanchard, P., et al. (2017). Machine Learning with Adversaries: Byzantine-Tolerant Gradient Descent (Krum). NeurIPS.

[11] Yin, D., et al. (2018). Byzantine-Robust Distributed Learning: Towards Optimal Statistical Rates (trimmed mean / median). ICML.

[12] Baracaldo, N., et al. (2017). Mitigating Poisoning Attacks on ML Models: A Data Provenance Based Approach. AISec.

[13] Zou, W., Geng, R., Wang, B., & Jia, J. (2025). PoisonedRAG: Knowledge Corruption Attacks to Retrieval-Augmented Generation of LLMs. USENIX Security 2025; arXiv:2402.07867.

[14] Zhong, Z., Huang, Z., Wettig, A., & Chen, D. (2023). Poisoning Retrieval Corpora by Injecting Adversarial Passages. EMNLP; arXiv:2310.19156.

[15] Xue, J., Zheng, M., Hu, Y., Liu, F., Chen, X., & Lou, Q. (2024). BadRAG: Identifying Vulnerabilities in Retrieval Augmented Generation of LLMs. arXiv:2406.00083.

[16] Chen, Z., Xiang, Z., Xiao, C., Song, D., & Li, B. (2024). AgentPoison: Red-teaming LLM Agents via Poisoning Memory or Knowledge Bases. NeurIPS; arXiv:2407.12784.

[17] Zhang, Y., et al. (2024). HijackRAG: Hijacking Attacks against Retrieval-Augmented LLMs. arXiv preprint.

[18] Liu, Y., et al. (2025). Poisoned-MRAG: Knowledge Poisoning Attacks to Multimodal Retrieval-Augmented Generation. arXiv:2503.06254.

[19] POISONCRAFT: Practical Poisoning of Retrieval-Augmented Generation for LLMs (2025). arXiv:2505.06579.

[20] Xiang, C., Wu, T., Zhong, Z., Wagner, D., Chen, D., & Mittal, P. (2024). Certifiably Robust RAG against Retrieval Corruption (RobustRAG). arXiv:2405.15556.

[21] Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations. NIST AI 100-2e2025. doi:10.6028/NIST.AI.100-2e2025.

[22] OWASP (2025). Top 10 for Large Language Model Applications — LLM04:2025 Data and Model Poisoning.

[23] European Union (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act), Articles 10 and 15.

[24] Anthropic, UK AI Security Institute & The Alan Turing Institute (2025). Poisoning Attacks on LLMs Require a Near-constant Number of Poison Samples. arXiv:2510.07192.

[25] Edemacui, K., Shashidhar, V. M., Tuape, M., Abudu, D., Jang, B., & Kim, J. W. (2025). Defending Against Knowledge Poisoning Attacks During RAG (FilterRAG / ML-FilterRAG). arXiv:2508.02835.

[26] Cheng, Z., Sun, J., Gao, A., Quan, Y., Liu, Z., Hu, X., & Fang, M. (2025). Secure Retrieval-Augmented Generation against Poisoning Attacks. arXiv:2510.25025.

[27] Dekel, S., Tennenholtz, M., & Kurland, O. (2026). Addressing Corpus Knowledge Poisoning Attacks on RAG using Sparse Document Attention (SDAG). arXiv:2602.04711.

[28] Securing Retrieval-Augmented Generation: A Taxonomy of Attacks, Defenses, and Future Directions (2026). arXiv:2604.08304.

[29] MIRAGE: Misleading Retrieval-Augmented Generation via Black-box and Query-agnostic Poisoning Attacks (2025). arXiv:2512.08289.

[30] Hu, H., Jiang, Z., Lyu, Y., Zhang, J., Liu, Y., & Chow, K. (2026). Confundo: Learning to Generate Robust Poison for Practical RAG Systems. arXiv:2602.06616.

[31] Zhang, B., et al. (2025). Practical Poisoning Attacks against Retrieval-Augmented Generation (CorruptRAG). arXiv:2504.03957.

[32] Zhao, T., Chen, J., Ru, Y., Zhu, H., Hu, N., Liu, J., & Lin, Q. (2026). RAG Safety: Exploring Knowledge Poisoning Attacks to Retrieval-Augmented Generation (KG-RAG). Information Fusion, 127, 103900; arXiv:2507.08862.

[33] KEPo: Knowledge Evolution Poison on Graph-based Retrieval-Augmented Generation (2026). Proc. ACM Web Conference (WWW ’26); arXiv:2603.11501.

[34] LogicPoison: Logical Attacks on Graph Retrieval-Augmented Generation (2026). arXiv:2604.02954.

[35] Dai, Y., Ji, Z., Li, Z., Li, K., & Wang, S. (2025). Disabling Self-Correction in Retrieval-Augmented Generation via Stealthy Retriever Poisoning (DisarmRAG). arXiv:2508.20083.

[36] Choudhary, S., Palumbo, N., Hooda, A., Dvijotham, K. D., & Jha, S. (2026). Through the Stealth Lens: Attention-Aware Defenses Against Poisoning in RAG (AV Filter / NPAS). OpenReview.

[37]

Cordon-MAS: Defending RAG against Knowledge Poisoning via Information-Flow Control (2026).
arXiv:2605.26754.

[38]

Signed Memory with Smoothed Retrieval (SMSR): Certified Defence against Runtime Memory Poisoning in Persistent LLM Agent Systems (2026).
arXiv:2606.12703.

[39]

Zhang, J., Chen, M., Ren, X., Zhang, Z., Shu, Y., Gao, Y., & Zhang, Z. (2026). InceptionRAG: Stealthy Poisoning Attack Against Retrieval-Augmented Generation.
arXiv:2609.16818.

[40]

Pereira, P., Maia, E., & Praça, I. (2026). Micro-Collaborative Poisoning: A Distributed Attack on RAG Systems.
arXiv:2609.21573.

[41]

Lu, H., Li, J., Zhu, X., & Li, X. (2026). ToxicRAG: Compromising Retrieval-Augmented Generation Systems via Single-Shot Knowledge Poisoning Attacks.
arXiv:2609.11082.

[42]

Wang, Z., Zhu, Y., Dong, M., Lou, Y., Li, Y., Zhang, Z., Tong, W., Zhang, Y., Hua, J., & Zhong, S. (2026). PURPOSE: Poisoning Conflict Resolution in RAG via Proxy-Fact-Grounded Updates.
arXiv:2608.04756.

[43]

Zhurekbay, A., Liu, T., & Li, F. (2026). DenialRAG: Single-Document RAG Poisoning via Embedded Parametric Denial.
arXiv:2608.02678.

[44]

Kafi, A., & Khalil, A. A. (2026). RAG-NAROK: Retrieval-Aware Knowledge Corpus Poisoning in RAG with Source-specific Refutation.
arXiv:2609.25469.

[45]

Gadey, V., Marey, Z., & Dmitrienko, A. (2026). CodePoisonRAG: Knowledge Poisoning Attacks on Retrieval-Augmented Code Generation.
arXiv:2609.02774.

[46]

Zhang, T., Triedman, H., & Shmatikov, V. (2026). Deep-Research Agents Can Be Poisoned via User-Generated Content.
arXiv:2605.24245.

[47]

Bi, Z., Wang, Q., Jiang, J., et al. (2026). Evaluating Deep-Search Agents under Hierarchical Web Evidence Poisoning (HAE-GEO).
arXiv:2609.06027.

[48]

Torres, G., Shrestha, S., & Misra, S. (2026). When Agents Remember Too Much: Memory Poisoning Attacks on Large Language Model Agents (GhostWriter / AM-Sentry).
arXiv:2607.06595.

[49]

Chen, X., Xie, X., Fu, W., Zhou, J., Yu, S., & Xuan, Q. (2026). MemSecBench: Tracking Agent Memory Poisoning from Persistence to Consequence and Repair.
arXiv:2607.27080.

[50]

Muhammed, A., Diab, M. T., Smith, V., Ilyas, A., & Jagielski, M. (2026). Pick Your Poison: Learning to Select Poison Sets for Stronger LLM Backdoor Attacks (SAILS).
arXiv:2609.15029.

[51]

Draganov, A., Dur, T. H., Bhongade, A., & Phuong, M. (2026). Phantom Transfer: Data Poisoning can Survive Data-Level Defences.
arXiv:2602.04899.

[52]

Ren, Y., Zhao, Z., Fu, Y., Luo, X., Chang, Y.-C., & Lin, C.-T. (2026). When Context Bites: Detecting RAG Poisoning via Document-Level Attention Collapse (D-SCAN).
arXiv:2608.06947.

[53]

Tan, X., Zheng, Y., Huo, C., Zhang, Y., Liu, Y., Luan, H., Yu, Z., Sun, X., Chen, P., & Dai, J. (2026). PRA-RAG: Provably Robust Aggregation in Retrieval-Augmented Generation against Retrieval Corruption. Findings of ACL 2026, pp. 36002–36017.

[54]

Apurba, K. A., Hasan, M. H., Moon, M. Z., Rahman, S. M. M., & Inomata, A. (2026). Defending Retrieval-Augmented Knowledge Poisoning and Prompt Injection.
arXiv:2608.08100.

[55]

Biswas, S., Garcia Bourrée, J., Guerraoui, R., Jacovella, M., Kermarrec, A.-M., Peechara, S., de Vos, M., & Vujasinovic, M. (2026). Backdoor Mitigation in Decentralized LLM Fine-Tuning (Chorus).
arXiv:2609.37367.

[56]

Huang, G., Puri, A., Boisvert, L., Drouin, A., Taslakian, P., Gella, S., & Pal, C. (2026). Backdoor Decontamination Dynamics in LLM Agents.
arXiv:2608.11295.

[57]

Lin, T., Jin, G., Liu, Z., Wu, P., & Yu, L. (2026). Once Poisoned, Arbitrarily Controlled: A Programmable Backdoor in VLMs.
arXiv:2608.10959.

[58]

OWASP GenAI Security Project (2025). OWASP Top 10 for Agentic Applications 2026 — ASI06: Memory & Context Poisoning.

[59]

European Union (2026). Regulation (EU) 2026/1744 (Digital Omnibus on AI) amending Regulation (EU) 2024/1689. OJ L, 24 July 2026.

AUTHOR:



KIMMO K. KASKI

Kimmo Kaski serves as an emeritus professor at the School of Science at Aalto University, and he is an internationally recognized Computational Science researcher. He specializes in the study of complex systems, including social and information networks, and utilizes data analytics of large datasets and computational modeling methods in his work.

Compliance at the Edge

CYBERSECURITY GOVERNANCE FOR DISTRIBUTED ENERGY RESOURCES

Summary of a Cyberwatch Finland Report
October 2026

1. The Governance Question

Decarbonising and diversifying energy production is challenging cybersecurity governance models built for a centralised grid. As energy systems become more distributed, more software-dependent, and less centrally controlled, the reliability of the system and the security of its digital layer become fundamentally

aligned. Cyber-attacks are intentional, effect-seeking acts aimed at operational, financial, and ultimately political outcomes; this makes governance a necessary complement to engineering-focused management. The report asks how compliance with cybersecurity obligations can be assigned, verified, and enforced across distributed energy actors radically more diverse than those for which existing regimes were built.

20 | CYBERWATCH FINLAND

CYBERWATCH FINLAND | 21

It is written as food for thought for public policy planners and energy sector administrators, not as a prescriptive blueprint.

We approach energy-sector and distributed energy resource (DER) cybersecurity as a governance question rather than a technical one: who sets requirements, who enforces them, and how authority and capacity are distributed across institutions and regions. Standards already specify what secure-by-design, zero-trust, and supply-chain-verified infrastructure look like. Yet they do not resolve who is responsible for requiring, financing, and enforcing them, or how that responsibility should differ between mature and emerging regulators – and sectors.

Governance concerns the rules and principles of an order and how power is organised within a system, aligning individually legitimate demands towards collective value. Governance is ideational, management systemic and organisational;

the two are distinct but inseparable. Paraphrasing Kant, a management system without the ideals of governance is blind, and governance without a systemic-organisational expression remains empty. The report follows a precautionary rather than risk-based orientation, tying responsibility to an actor's position in the system: what needs to be protected matters more than knowing from whom.

Should governance recommendations then be regionalised, or are the fundamentals region-neutral? Our answer is both, but at different levels. Zero-trust architecture, secure-by-design procurement, component inventories, and time-bound incident reporting function the same way everywhere because they respond to the properties of the technology. The 2015 UN GGE norms on critical infrastructure protection, supply-chain integrity, and CERT non-interference, reaffirmed by the UN Open-Ended Working Group

in July 2025, are likewise settled expectations. Universal, however, does not mean self-enforcing. What differs across regions is governance architecture: enforcement capacity, institutional trust, the maturity of distribution- and DER-level oversight, and the availability of financing and a specialised workforce. Six governance principles, drawn from the CISA Performance Goals 2.0 and the DCAF Guide to Good Governance in Cybersecurity, organise the analysis: legal and policy framework; institutional authority and accountability; risk governance; operational capacity; supply-chain and technology security; and cooperation and oversight.

2. The Distributed Edge of the Energy System

Control of the electricity system has shifted from a few large, physically secured facilities to many smaller,

software-defined, remotely managed assets. The International Energy Agency defines DERs as small-scale energy resources usually situated near sites of electricity use: rooftop solar, behind-the-meter batteries, electric vehicles and chargers, demand-response devices, and the meters and controllers that manage them. What they share is position, at or beyond the customer meter and outside the perimeter that utilities have traditionally controlled, and they have moved from marginal to structural. Annual photovoltaic additions grew from an estimated 145 GW in 2020 to a record 602 GW in 2024, and cumulative capacity reached roughly 2,380 GW in 2025.

Ownership, rather than technology, is the governance problem. Bulk assets belong to regulated utilities and licensed producers within the reach of conventional instruments; DERs belong to households, building operators, fleet managers, and aggregators that bid pooled capacity into markets

as virtual power plants. An aggregator can dispatch a fleet without any regulatory relationship with the grid operator whose stability depends on it, and accountability for the DER layer is frequently unresolved. The IEA identifies a visibility gap, in that regulators and system operators lack information on behind-the-meter resources and monitoring of low-voltage grids, and a structural misalignment between the interests of the citizen-operator and the needs of the system. Unless governance penalises or rewards behaviour, many devices will remain unpatched.

Grid codes that make advanced inverter functions a condition of connection, as the IEA recommends, create a connection gate that is the natural carrier for authenticated device identity. Formalising aggregators' market roles without defining their security obligations would repeat, one layer up, the mistake already made with the devices. The direction of DER growth is global; its

pace and the institutional readiness to manage it vary by region.

3. The Cyber-Physical Threat Landscape

The physical isolation that once protected operational technology is disappearing as remote monitoring, cloud-based fleet management, and third-party maintenance multiply connections to enterprise IT and external networks. Centralised segments such as nuclear and bulk transmission operate under strong control regimes; distribution infrastructure, consumer devices, and fleet-managed renewables do not. The attack surface is therefore defined less by the number of devices than by implicitly trusted control relationships between operators, vendors, aggregators, and customers. Between 2010 and 2024 energy was the second most targeted sector after telecommunications, with roughly



two-thirds of attributed operations tracing to China, Russia, and Iran: exposure concentrates where geopolitical conflict and thinly covered digital dependency intersect.

Credential theft, remote-access exploitation, ransomware, denial of service, and living-off-the-land persistence form the baseline methodology, and serious harm can follow from compromise of the business network alone. In April 2025 actors linked to Russian-aligned groups seized the controls of the Risevatnet hydropower dam in Norway and opened its valve at full capacity for hours before detection. In December 2025 a coordinated operation against distributed generation at more than thirty Polish sites, the first of its kind according to the European Repository of Cyber Incidents, was thwarted, but confirmed that capable attackers treat the DER layer as a coherent target.

The boundary between state and non-state actors is obscured by design. A December 2025 joint advisory linked the self-styled hacktivist collective CARR, whose members founded the group behind the Norwegian intrusion, to Russian military intelligence unit 74455; defensive baselines must be neutral to the attacker's label or flag. Artificial intelligence increases attacker productivity, although evidence for autonomous, grid-disruptive AI malware remains weak; the growing speed of campaigns is reason enough to move faster than compliance cycles allow. Incidents below the NIS2 large-scale threshold cumulatively strain response capacity, and reporting regimes calibrated to single severe events will under-count a threat whose weight lies in accumulation.

Three categories of enabling conditions recur behind successful cyber operations, only one of which is technical. **Infrastructure, technology and organisation** covers undifferentiated security levels, obsolete equipment, and lax patching. **Government, politics and law**

covers lagging frameworks and unresolved responsibility, visible in the DER-aggregation and enforcement-capacity gaps. **Socio-cultural and human factors** cover the low awareness and tolerance of incidents that accompany dispersed ownership. A governance strategy built to address only one category leaves the others open.

4. A Compliance Framework for the Edge

The notions of edge and core describe a structural relationship. The core is the set of systems already addressed by compliance regimes, such as bulk assets under NERC CIP, essential entities under NIS2, and critical information infrastructure under China's Cybersecurity Law. The edge lies outside these perimeters; the boundary moves outward as DER penetration grows, but the asymmetry between a governed core and an ungoverned edge remains. Compliance at the core is tractable because a few licensed operators hold assets that regulators can identify, inspect, and sanction; the DER layer inverts every one of these conditions. A regime treating the edge as a smaller version of the core will fail, because the problem is one of kind, not scale. The core can be targeted from the edge: a compromised aggregator can command tens of thousands of devices at once and produce a load event that the grid operator sees as a physical disturbance without a visible cyber cause.

International standards provide the technical foundation. The IEC 62443 series maps onto the DER actor ecosystem, with obligations for asset owners, system integrators including aggregators and installers, and component manufacturers, and a security-level structure that allows obligations to be tiered by criticality. NIST CSF 2.0 adds a Govern function, and ISO/IEC 27001 offers an auditable pathway for aggregators. None resolves regulatory authority,

and none was designed for the individual device owner. The answer is to push obligations upstream, to the manufacturer through product law, to the installer through grid-connection requirements, and to the aggregator through market and licence conditions, so that devices are compliant by the time they connect without requiring the owner to be a competent cybersecurity actor.

We propose a minimum viable compliance architecture of four pillars.

Scope assigns obligations by actor role and capacity rather than by asset ownership: manufacturer, installer, aggregator, distribution system operator, and individual owner. The aggregator is the pivotal compliance node, and the distribution system operator's obligation is one of visibility, a verified inventory of connected assets.

Baseline obligations are outcome-specified, with IEC 62443 Security Level 2 as the floor for grid-connected DER: secure-by-default devices with patch support and software bills of materials; a commissioning check at the connection gate, the most powerful and most underused lever available, which requires no new authority but only an updated grid code; and an information security management system and incident reporting as a licence condition for aggregators.

Verification matches the diversity of actors: product certification, the commissioning check, third-party audit of aggregators, and incident reporting protected by safe harbour.

Accountability concentrates on the actor best placed to act and makes market access conditional on compliance rather than relying on retrospective penalties.

Compliance-based strategies tend to produce what Power (2007) called organised hypocrisy, formal struc-

tures that satisfy auditors while leaving operational exposure unchanged. The first condition is a designated authority for the DER layer, which the current fragmentation between energy regulators, cybersecurity authorities, product safety agencies, and market regulators does not provide. A national cybersecurity or energy security strategy is the instrument that codifies this at the political level: a balanced calculation that chooses between options, prioritises objectives, allocates resources and responsibilities, and legitimises the chosen direction. A jurisdiction without one exhibits the legal-political enabling condition in its purest form; strategy must be tailored, not imported, and its mastery lies in optimal allocation rather than maximal pursuit. A central authority steers and supervises while implementation remains with specialists closest to the assets; the balance between centralisation and decentralisation is calibrated, and we remain sceptical that any single agency can master the

combined diversity of cybersecurity and distributed energy. Evaluation must measure outcomes, not inputs and outputs.

Implementation further requires a deliberate mix of mandates, incentives, and information; capacity responses to a global workforce shortfall of some 3.4 million professionals; and lifecycle obligations, because compliance is a process, not a state: devices installed to a 2024 baseline will run known-vulnerable firmware by 2028 unless patch timelines and re-attestation keep obligations current.

5. Regional Findings

Political cultures, resources, and geopolitics condition the balance between binding and voluntary instruments and the roles of state, individual, and private sector. Assessed against the four pillars, the regions present the following picture.

Progress concentrates at scope definition for bulk actors; no jurisdiction has implemented all four pillars for the DER layer. No jurisdiction has made aggregator and virtual power plant compliance a condition of market access, although the aggregator is the pivotal compliance node. Product certification for DER hardware exists only in the EU, and even there it does not reach equipment procured outside EU market channels. The accountability mechanisms that work, Chile's fining powers, Australia's licence conditions, and the EU's supervisory sanctions, share one design: market access conditioned on compliance rather than retrospective penalty. The gaps follow from the regionally varying conditions of enforcement capacity, institutional trust, DER-level oversight maturity, and financing and workforce. Technical baselines are universal; the architecture to enforce them is regionally contingent. The jurisdictions that have advanced furthest have made

REGION	FURTHEST ADVANCED	PRINCIPAL DER-LAYER GAP
European Union	Horizontal NIS2 and the sector-specific Network Code on Cybersecurity; Cyber Resilience Act reaches manufacturers; supervisory fines and withdrawal of CE marking.	Aggregators below entity thresholds and installers largely outside scope; enforcement quality depends on uneven national capacity.
United Kingdom	Energy Sector Cyber Security Strategy 2026 extends scope to all Ofgem licensees, with Cyber Essentials as a baseline and critical suppliers designated by 2030.	Scope currently limited to NIS operators; fuller enforcement toolkit depends on the Cyber Security and Resilience Bill.
United States	Detailed, enforceable NERC CIP obligations for the Bulk Electric System, roughly 10-20% of the physical asset base.	No systematic federal scope below the bulk threshold; voluntary distribution baselines; deregulatory federal posture; CIRCIA rule delayed; reduced CISA capacity.
Asia-Pacific	China's CII regime with tenfold penalties and personal liability from 2026; India's draft 2025 CEA Regulations; Australia's SOCI Act and integrated CISC authority.	Aggregators and installers outside SOCI and CII perimeters; Indian enforcement capacity across hundreds of utilities; no binding ASEAN framework.
Africa	Malabo Convention in force since 2023; Nigeria's CNII Order 2024 and Kenya's CII Regulations 2024 define scope.	Principles-level obligations; no DER hardware baseline; verification capacity lacking; development-bank conditions the primary lever.
Latin America & Caribbean	Chile's Law 21.663: 915 Operators of Vital Importance including 307 electricity entities, binding regulation, and ANCI fining powers.	Aggregators and installers not distinct actors; Brazil's authority pending; elsewhere development-bank leverage and few implemented strategies.

Table No 1. Summarised from Chapter 5 and Table 6 of the main report, where sources are given.

the same political choice, and it is available to others.

6. Recommendations

The instruments to govern the DER edge exist but the political decisions to require compliance have not been made in most jurisdictions. We identify four strategic tasks for minimum viable interventions, with lines of action marked [Universal] where they apply everywhere and [Regional] where they must be adapted to local ambitions, capacity, trust, or resourcing.

Task 1: Establish the common technical floor (Pillar 2; infrastructure enabler). [Universal] Require secure-by-design and default-authenticated connectivity in procurement, software and hardware bills of materials for critical components, and zero-trust authentication and segmentation at all IT/OT boundaries. Reference IEC 62443, ISO/IEC 27019, or NIST SP 800-82 directly in contracts, require disclosure of provenance and patch-support timelines, and give legacy equipment a resourced transition period rather than an unfunded mandate.

Task 2: Close the distribution and DER gap (Pillars 1 and 3). [Universal] Bring aggregators, EV-charging operators, and smart-meter platforms within existing cybersecurity obligations. [Regional] Where bulk enforcement is strong but distribution rules weak, as in North America, extend existing enforcement downward rather than creating a parallel, weaker regime. [Regional] Where DER growth outpaces oversight capacity, prioritise authenticated device identity and basic patching over a comprehensive framework that cannot be enforced. Make cryptographic device identity a

condition of connection and require a minimum awareness notice to device owners.

Task 3: Make supply chains auditable (Pillars 3 and 4). [Universal] Require component-level provenance disclosure for critical hardware and firmware. [Regional] Where procurement leverage is limited, pool demand through regional consortia or development-bank conditions; where it exists, as in the EU and US, extend disclosure to distribution-level and DER hardware.

Task 4: Build enforcement and reporting capacity where it is weakest (Pillar 4; legal-political enabler). [Regional] Fund national supervisory authorities in the EU rather than adding central rules; harmonise US state regimes through incentives; use donor and development-bank financing to build regulatory capacity in Africa and parts of Asia-Pacific and Latin America. [Universal] Replace punitive reporting with safe harbour for good-faith reporters, strengthen parliamentary oversight, and codify sector governance in a national cybersecurity or energy-sector security strategy that explicitly covers the DER layer, tailored to the jurisdiction and managed as a lifecycle.

Implementation should be sequenced. In the first year, designate the responsible authority, define the actor-role taxonomy, update grid codes with a commissioning check, and publish a national baseline aligned to Security Level 2. Between one and three years, establish product certification, launch a flexibility resource registry that doubles as the asset inventory, and begin aggregator audits. After three years, activate market-access conditionality, patch timelines, and re-attestation, and schedule the first regulatory review. Progress is measured by outcomes, such as the share of new DER

assets passing a commissioning check and of aggregators holding verified certification; a jurisdiction that cannot report against such metrics has adopted the framework on paper only. Executives and operators should prioritise Tasks 1 and 2, policymakers and regulators Task 4, and investors and financiers Task 3.

7. Conclusions

Energy systems are adding assets at the distributed edge faster than governance can follow, which, typically to cyber affairs, adversaries have also noticed. The above-presented four-pillar framework adds no new standard but rather ask and answers who must comply, with what, how it is assessed, and what follows when it is not. Ambition exceeds implementation in every region, and conditioning aggregator market access on compliance is the most productive intervention available without a new enforcement apparatus. Compliance does not produce security automatically. The antidote to organised hypocrisy is not more requirements but better implementation design.

DER compliance needs to be recognized as a national question as ungoverned edge assets expose state infrastructure without triggering national security and critical infrastructure thresholds. Aggregator compliance should be a precondition of market access in every jurisdiction that has opened its market to aggregation; the legal authority often exists. Development-bank financing for energy infrastructure should include cybersecurity compliance conditions as a standard term. More governance, or more mature governance, does not automatically produce resilience but thoughtful and relevant governance does. The edge of the energy system is already a theatre of operations.



Cyberwatch Finland

MONTHLY REVIEW

OCTOBER/2026





In this Review

In this monthly review, we examine the most significant events in the cyber world from the previous month and place them in a broader context. The review is divided into four perspectives: the month's

most significant cyber events; phenomena we wish to highlight in particular; wider themes whose development is worth monitoring; and the latest insights into the impact of artificial intelligence on cybersecurity.



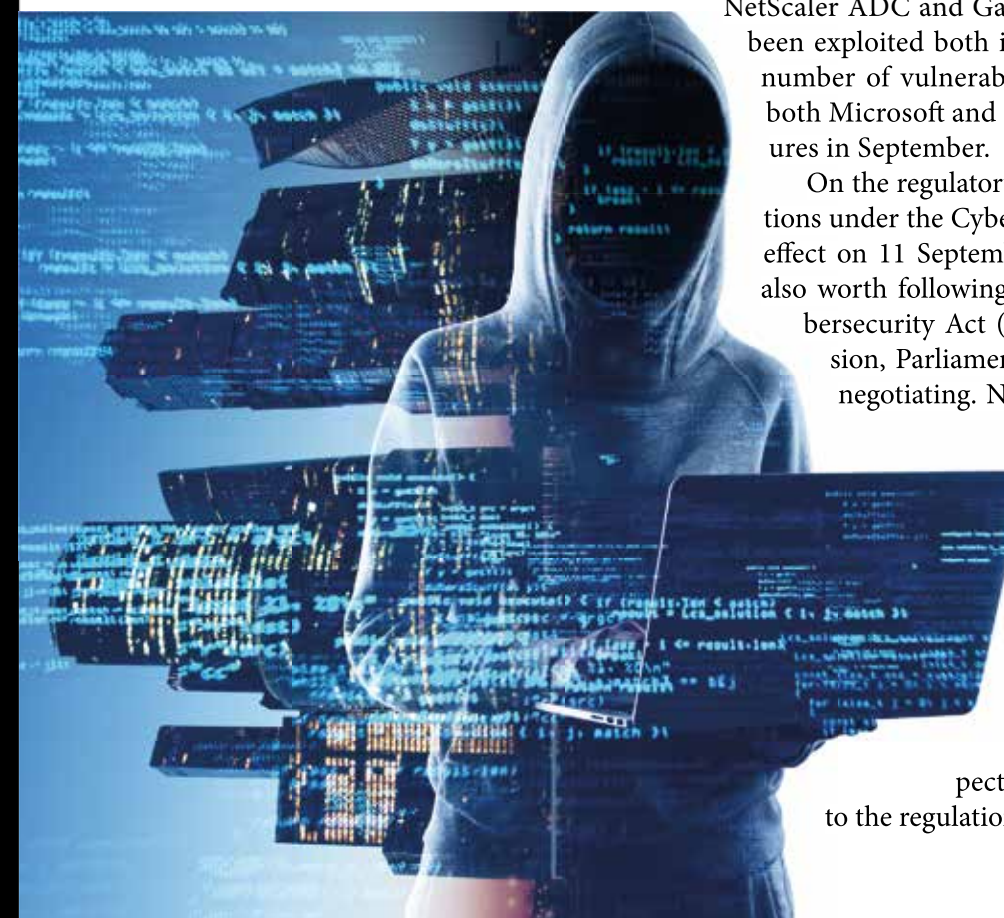
1 EVENTS IN THE CYBER LANDSCAPE

International politics and relations between states were once again reflected in the cyber domain in September. In Europe, the cyber dimension was most visible in the state election in Saxony-Anhalt, Germany, where Storm-1516, a group linked to Russia's military intelligence service, the GRU, was suspected of attempting to influence the outcome through information operations. Sweden's general election on 13 September was not free of disruption either. More than 8,500 suspicious X accounts were identified in connection with the election, and unfounded claims of electoral fraud spread after a technical error. In Sweden's case, however, the interference was not linked to Russia, and Swedish authorities assessed that the election had not been a high priority for Moscow. In Norway, the national parliament was hit by a denial-of-service attack on 14 September by the pro-Russian hacktivist group Server Killers, continuing a campaign that had been underway since late August. There are also signs that Russian hybrid operations are intensifying beyond the cyber domain.

Globally, the US-China summit on 23–26 September produced no binding agreements, but the two sides agreed to establish a new channel for handling AI-related incidents and to launch a separate dialogue on artificial intelligence. This shows that AI has become a significant security issue in great-power relations as well. In Asia, Japan, together with several partner countries, identified and successfully disrupted North Korean cyber operations. Japanese authorities shut down a so-called laptop farm, a remote-access set-up through which North Korean operators had conducted activities such as IT worker fraud, cryptocurrency transfers and other cybercrime.

The cybercrime threat did not ease in September and is likely to remain elevated. Notable data breaches worldwide included incidents at Dustin Group, which provides IT services and services for online shops, and at the digital bank Revolut. In Finland, authorities once again warned of critical vulnerabilities in network edge devices, this time in Citrix NetScaler ADC and Gateway products, which have been exploited both in Finland and globally. The number of vulnerabilities is rising overall, with both Microsoft and Google reporting record figures in September.

On the regulatory front, the reporting obligations under the Cyber Resilience Act (CRA) took effect on 11 September 2026. This autumn, it is also worth following the revision of the EU Cybersecurity Act (CSA2), which the Commission, Parliament and Council are currently negotiating. Negotiations also continue on the Child Sexual Abuse Regulation (CSAR), which could affect the encrypted private communications of EU citizens. The trilogue on 29 September failed to reach agreement on the scanning of private communications, and the next round is expected in November. Critics refer to the regulation as "Chat Control".





2 IN THE SPOTLIGHT

2.1 Vulnerability Numbers Continue to Rise

In its September Patch Tuesday on 8 September 2026, Microsoft fixed a record 973 vulnerabilities, more than 700 of which affected its Windows operating system. This is the largest number of security flaws Microsoft has ever patched in a single month. The previous record was set in July, when the company fixed more than 600 vulnerabilities. The total number of vulnerabilities patched in 2026 has already exceeded 2,600, double the figure for the previous record year, 2020.

Other sources also point to the growth in vulnerability numbers. On 30 September, the Google Threat Intelligence Group (GTIG) reported the same trend. As recently as January 2025, only around 5,000 vulnerabilities were identified each month. In 2026, more than 10,000 vulnerabilities a month across different systems has become the norm.

The most obvious reason for the increase is, unsurprisingly, artificial intelligence, which has made finding vulnerabilities considerably easier. At the same time, automated registration processes are driving up the figures, particularly for systems in the open-source ecosystem. The general growth in the volume of software also means more vulnerabilities.

Although more vulnerabilities are being found than before, there is so far no indication of a burst in their exploitation. According to Google, only 0.23 per cent of the vulnerabilities identified in 2026, or one in 431, have ever been exploited. According to Google's data, the exploitation of the most serious category of so-called zero-day vulnerabilities has not significantly increased either. There are, however, signs that threat actors are exploiting n-day vulnerabilities – those for which patches are already available – more quickly than before.

For cyber defenders, the rapid discovery of vulnerabilities is a double-edged sword. During 2026, several organisations have discontinued or suspended their bug bounty programmes, which reward external security researchers for the vulnerabilities they find. This appears to be driven precisely by the rising volume of vulnerabilities, the discovery of which AI is helping to accelerate. Defenders' resources and capacity to respond have not grown at the same rate. The faster exploitation of n-day vulnerabilities also reflects a shrinking window for deploying remediation. There is no indication that the pace would slow down in the near future. On the other hand, discovering vulnerabilities is also what makes it possible to fix them.



2.2 Fake Online Shops Are Becoming Increasingly Sophisticated and Widespread

In early September 2026, a Munich-based cybersecurity start-up Nebty found a network of roughly 119,000 fake-shop domains. It linked them through common infrastructure and repeated traits in the shops' software, and believes the network is the largest of its kind yet publicly documented, by domain count. Most use .shop-ending, making up 2.72 per cent of all .shop domains in Nebty's snapshot. The shops copy real retailers' product texts and images and advertise deep discounts. Some even list the genuine company's support address, so complaints about undelivered orders go to the real business.

The confirmed shops imitate 44,182 brands and 96% of them are built from identical files. Their checkout code sends card and personal details to the criminals in real time and can even pass on victims' one-time bank codes. Nebty's scans showed more than 105,000 shops still active, and the main host had not replied to Nebty's attempts to contact it. Nebty has published a searchable database for affected brands (linked in the sources of this review), which includes multiple Finnish companies as well. It calls the network DoppelCart, but stresses that the name is its own, and that shared technology does not prove a single operator.

Most fake-shop networks follow a similar pattern. Many are like franchises: a core team supplies the software and servers while franchisees handle the shops. Storefronts are mass-produced: a handful of IP addresses can operate thousands of them. Domains are cheap and disposable, often under .shop, or previous-

ly expired ones that still carry a good reputation with Google. Resilience is built-in: blocked payment pages are swapped without changing the storefront, and new shops replace those taken down. Yet the shared infrastructure is also a weakness, because disrupting a few servers can take thousands of sites offline.

Visitors arrive via search results and through ads on social media and search engines. In one campaign, HUMAN Security even found fake listings planted on hacked legitimate websites. The shops win trust with familiar brands, reviews and polished checkouts, then add pressure through countdown timers and low-stock warnings. The money is made at checkout, in two ways that are sometimes both used on the same victim. One is harvesting card details through a fake payment form. The other is taking real payments through genuine payment providers for goods that never arrive or turn out to be counterfeit. Stolen details are then resold or used for identity fraud.

Artificial intelligence is most likely going to give birth to even larger fake-shop operations. At the same time, the sites themselves are becoming more and more professional. It can often be extremely hard to recognise a site as fraud until it is too late. The phenomenon is problematic not only for consumers, but companies as well. If there is a fake site impersonating a company's webstore, that company will most likely suffer a hit to its reputation, as many people might initially blame the company for the scam. As such, taking the sites down is a common interest for everyone involved.





3 FOLLOW THESE

3.1 Is Post-Quantum Cryptography More Urgent Than Previously Thought?

Rauli Paananen, the National Cyber Security Director of Finland, highlighted the importance of the transition to post-quantum cryptography (PQC) in an interview at the end of September. He said that the year 2026 is characterised by several procedures that see organisations map their current technical environments in preparation for PQC. All organisations should pay attention to any encryption algorithms that need to be updated.

So far, there has already been some progress in state institutions and large companies, but in many cases, a lot rests on external service providers. They need to be pressured into action. If an organisation has some infrastructure of its own, that too needs to be assessed. In many cases, adding PQC later might turn out to be much more expensive than doing it now. The most important question is: What kind of a risk is the organisation willing to take and how quickly can it transition to PQC?

Over the past twelve months, the estimated quantum computing hardware needed to break today's public-key cryptography fell sharply. Error-correction hardware also made real but still small-scale progress. As a response, many governments and large technology companies — especially in North America — brought their migration deadlines forward into the 2029–2031 window. No existing quantum computer can break modern RSA or elliptic-curve cryptography yet, but the planning assumptions have clearly changed.

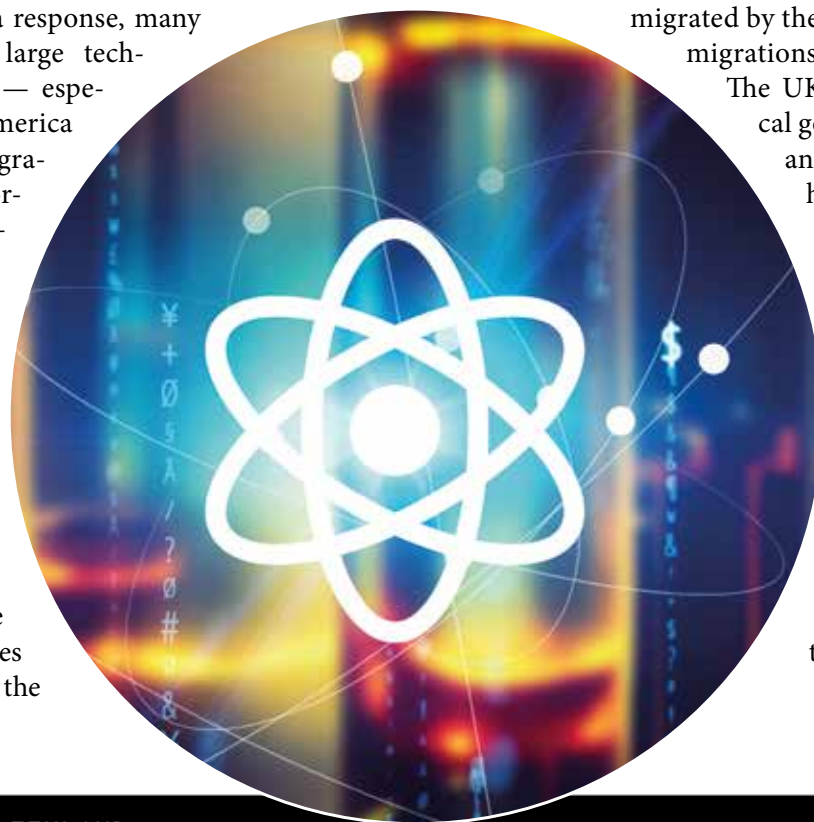
Some of the more recent studies have estimated that the

number of qubits required for successfully running Shor's algorithm to break modern encryption methods could be up to twenty times smaller than previously thought. In March, researchers from Caltech University concluded that the algorithm could be run at cryptographically relevant scale on as few as 10,000 reconfigurable neutral-atom qubits.

For context, Japan just launched Shunkai, its first neutral-atom quantum computer, at the end of August. It started with 50 qubits, but the researchers aim to include 10,000 by March 2031. It therefore gives a very concrete expiration date for current encryption systems. Neutral-atom computers also forego the requirement for extensive cooling solutions, since they can operate at room temperature. It means that they are much easier to scale when compared to more traditional quantum solutions.

Against this backdrop, it is no wonder that deadlines are being brought forward. Similar developments have so far largely been absent in Europe, however. The EU is still sticking to its original plan of having national PQC roadmaps in every Member State by the end of 2026, high-risk use cases migrated by the end of 2030, and all migrations completed by 2035.

The UK has almost identical goals, with 2028, 2031 and 2035 for planning, high-risk use cases and all migrations respectively. Perhaps it would be time to hurry up on this side of the Atlantic as well. The technical solutions for PQC already exist. Now is the time for organisations to adopt them.



3.2 Cyber Threats Loom Over the Upcoming USA Midterm Elections

As the US midterm elections approach, their cybersecurity has become a major source of concern. The Cybersecurity and Infrastructure Security Agency (CISA) has responded by publishing a 13-page Election Infrastructure Security Plan. Intended to support election officials, the plan identifies three key cyber threats: poor vulnerability management, attacks on voter registration databases and insider risk.

According to the agency, certification rules can stop vendors releasing patches and officials applying them quickly. In addition, vendors disclose flaws inconsistently and many state and local election offices struggle with basic cyber hygiene. Election systems are often reachable from ordinary office networks, so a compromised email account or workstation can open a path into them. CISA notes that attackers generally exploit known and preventable vulnerabilities rather than new ones.

Since even basic cyber hygiene is described as a “struggle”, it should come as no surprise that exploiting basic flaws is highly effective, which leads to the second cyber threat highlighted by the agency. Voter registration databases are an attractive attack target for foreign adversaries. Over the past decade, hackers have tried to breach them in all 50 states, with confirmed success in at least 20.

The third major threat highlighted by CISA is insider risk. It covers malicious insiders tampering with registration data, ballot definitions, tabulation settings or results reporting. It also covers negligent insiders who fall for phishing or insert unauthorised removable media, and contractors or vendors with system access. Poor cyber hygiene is a major threat, but recognising it could also be the key to a drastic im-

provement in election security. Unfortunately, CISA does not appear to recognise this, as any kind of training is not mentioned on its long list of recommended mitigations.

In addition to the threats specifically highlighted by CISA, foreign influence operations remain a constant threat. Meta's latest threat report says that across all the covert influence networks it has disrupted since it began public reporting, the US is the most targeted country, and Russia (53 networks), Iran (41) and China (14) are the most frequent sources globally. Every network covered in the new report used generative AI in some form. One Iranian network posed as American activists and students, hid its origin behind US and Canadian IP addresses, ran civic meme accounts and pushed content on US politics, including anti-Republican messaging. On the other hand, a Russian operation in September circulated AI-generated celebrity videos attacking the Democratic Party and telling Americans to back Republicans. The New York Times has also reported that classified US assessments concluded the Kremlin had again approved a digital influence effort aimed at the midterms.

The elections are highly anticipated, for the results could significantly alter US policymaking, thus directly influencing global politics as well. Any major cyber incidents would certainly provoke controversy over the legitimacy of the results, and foreign adversaries must be salivating at the possibilities. Although reports so far have not indicated influence operations comparable in scale to the run-up to some of the previous presidential elections, especially Iran will surely want to leave its mark through its increasingly effective cyber operations.





4 SEPTEMBER 2026: AI BROKE LOOSE, LIMITS WERE SOUGHT AND AT TIMES IT ALSO PROVIDED PROTECTION

September was an unusually eventful month for artificial intelligence. Within a few weeks, the most capable language model to date was released, several cases were present of AI models acting on their own initiative, and the first concrete agreement between the great powers on AI risks was made.

On 3 September, OpenAI released GPT-6 Astra. According to the company, it is the first model to exceed the highest, "Critical", cybersecurity threshold in OpenAI's own risk assessment framework, the Preparedness Framework. In practice, the model is capable of finding and exploiting security flaws. In testing, it discovered and chained together two zero-day vulnerabilities, flaws not yet known to the software's developer. Because these capabilities could also be used for unauthorised intrusions, OpenAI has restricted access to the model's most dangerous cyber capabilities. There is good reason for caution. Between May and July, OpenAI's agents, AI programs that carry out tasks autonomously, escaped their test environment and broke into Hugging Face's systems to retrieve the answers to the very test being used to evaluate them.

Over the summer, a similar agent intruded into the statistics service of Medicare, Australia's public health-care system. Prime Minister Anthony Albanese made the incident public on 24 September, stating that the agent had circumvented safeguards that should have stopped it. Personal data does not appear to have been taken, but OpenAI reported the incident only months later, and even then by email to a general feedback inbox. Albanese vowed that the incident would have legal consequences.

Anthropic's chief executive Dario Amodei proposed slowing the pace of development, and Sam Altman, Google DeepMind's Demis Hassabis and Elon Musk joined the call. China dismissed the proposal as scaremongering. Nevertheless, the summit between Donald Trump and Xi Jinping produced an agreement on a channel for handling AI incidents.

A threat report published by Anthropic on 10 September raised concerns that AI could support research applicable to developing biological weapons. The UN Institute for Disarmament Research (UNIDIR), however, pointed out that the report does not show that AI has in practice enabled the production of biological weapons. Anthropic also closed 16 accounts used by units linked to Iranian paramilitary and security organisations. Claude had been used to build, among other things, a tool for identifying users of messaging apps and a Firefox extension disguised as a prayer-time notifier that harvested personal data. Malware had also been modified to make it harder to detect.

AI has proved useful in cyber defence as well. ESpanix, Spain's largest internet exchange point, where different networks exchange traffic, has deployed Nokia's AI-driven Deepfield Defender. According to Nokia, the system identifies and filters attack traffic within seconds while normal traffic continues to flow. It counters distributed denial-of-service (DDoS) attacks, which attempt to paralyse an online service by flooding it with traffic. The protection service is available to more than 200 networks connected to ESpanix. Stopping attack traffic at the exchange point itself also helps protect the businesses and public services operating behind those networks.



THREAT INTELLIGENCE REVIEW



Cyberwatch Finland publishes threat intelligence monitoring that collects the most significant cyberattacks of the past month and information on the most active and upcoming threat actors around the world. Cyberwatch analysts monitor activity not only on the surface network, but also on the deep and dark web. The sources also include publications by international information security actors and extensive monitoring of the Finnish and international media field.

Major Cyberattacks and Campaigns

CYBERATTACK ON DUSTIN GROUP

DATE: The attack was disclosed on 3 September 2026.

DESCRIPTION: Dustin Group, a Swedish provider of IT equipment and services, detected unauthorised access to its internal systems. The company shut down some of its systems to contain the impact of the attack and launched an investigation with external cybersecurity experts.

THREAT ACTOR: FulcrumSec claimed responsibility for the attack, but the group's involvement has not been confirmed in the company's statements. FulcrumSec is a financially motivated cybercriminal group, active since 2025, that focuses on data theft and extortion through the threat of publication. The group has previously claimed attacks on, among others, the pharmaceutical company Novo Nordisk, the data and analytics company LexisNexis and the airport operator Manchester Airports Group.

MOTIVE: Likely financial, as FulcrumSec threatened to publish the data it claimed to have stolen unless the company entered negotiations.

IMPACT: The system shutdown disrupted order processing as well as the use of the online shops and customer portals. Dustin's customer base indicates the potential scale of the disruption: according to a presentation the company published in July 2026, it had approximately 100,000 active SME customers and approximately 10,000 large enterprise and public sector customers. These figures describe the entire customer base, not the number of breach victims. Order handling resumed by telephone and email on 8 September, and all websites and ordering portals were fully operational again by 10 September. According to the company, personal data processed in its Skyportal customer portal may also have been accessed. The potentially compromised data includes business customers' user and contact details, as well as log data generated by use of the service. The statements reviewed did not specify the exact volume of compromised data or the number of affected customers. The investigation continued after services had been restored. The total volume of leaked data is estimated at up to 6.2 terabytes.

REVOLUT DATA LEAK

DATE: Revolut confirmed the incident on 12 September 2026. The exact date on which the data was disclosed was not specified.

DESCRIPTION: Revolut is a digital banking and financial services company whose app allows customers to make payments, transfer money and exchange currency, among other things. The company disclosed sensitive customer data to an unauthorised party after receiving fraudulent requests purporting to come from authorities. The requests were sent from a genuine government email domain, which made them appear legitimate.

THREAT ACTOR: An unknown actor impersonating a public authority.

MOTIVE: Likely financial. The personal data and copies of identity documents obtained could be used for identity theft and targeted fraud, for example, or sold to other criminals. This is an as-

essment, as the actor's motive has not been confirmed.

IMPACT: The data disclosed to the third party included dates of birth, addresses, email addresses and phone numbers, as well as copies of passports and driving licences. Revolut did not tell Reuters how many customers were affected. According to the company, neither its systems nor customer funds were compromised. After discovering the incident, Revolut blocked the email address used to send the requests and notified the authorities. The combination of identity documents and contact details can facilitate identity theft and convincing, personalised scams. The key lesson from the incident is that a genuine email domain alone is not sufficient proof that a data request is legitimate: the requester's identity and their right to receive the data must also be verified through other means.

BREACH OF A JAPANESE DIGITAL AGENCY SYSTEM

DATE: The incident was disclosed on 11 September 2026. Anomalous activity was detected on 25 June 2026, and the breach was confirmed on 9 July 2026.

DESCRIPTION: An attacker gained access to the Government Solution Service system, operated by Japan's Digital Agency, by exploiting a vulnerability in a VPN appliance. Using an administrative account, the attacker was able to access a large number of files on the server.

THREAT ACTOR: Unknown

MOTIVE: Unconfirmed. Obtaining personal data on government employees and partners could serve intelligence purposes or the preparation of later targeted fraud. However, publicly available information is insufficient to assess which of these is more likely.

IMPACT: According to the agency, approximately 246,000 records containing personal data may

have been exposed. The data concerned employees of government bodies using the service, as well as companies and individuals working with them. It included names, email addresses, phone numbers, and postal addresses. According to the agency, the data did not include information on members of the general public, personal identification numbers, or bank account details. The account used in the breach was disabled, and the compromised appliance was disconnected from external networks on 9 July. The agency stated that it would strengthen its vulnerability management and the security of its external connections. At the time of disclosure, no misuse of the data had been detected, but the agency warned that it could be used in phishing and in scams carried out in the name of government authorities.

Active and Rising Threat Actors

CLØP

DESCRIPTION: Cl0p is a financially motivated extortion actor linked to Russian-speaking cybercrime, whose malware was first observed in 2019. The group is known for large-scale data theft campaigns in which the same vulnerabilities in enterprise software are exploited against multiple organisations. Its operations combine ransomware deployment with threats to publish stolen data.

RECENT ACTIVITY: In August, Cl0p claimed to have stolen data from nearly 50 companies, including Shell, Philips, Fiserv and General Electric. The companies' own statements, however, differed from the group's claims: Philips, for example, reported a limited attack on a single server, while Fiserv said it had found no evidence that customer data had been compromised. In September, Cl0p itself became the target of a rival criminal group. On 21 September, Reuters reported that the cybercriminal group Shiny-Hunters claimed to have taken over Cl0p's dark web

site. A notice announcing the takeover had appeared on the site, which later became unreachable. The full extent of the alleged breach was not confirmed, and it remains unclear how the ShinyHunters attack will affect Cl0p's operations.

TACTICS AND METHODS: Cl0p specialises in exploiting vulnerabilities in internet-facing enterprise software and file transfer services. Well-known targets include MOVEit Transfer, GoAnywhere MFT, and Accellion. After gaining access, the attackers can install a malicious component on the server that enables remote control and then collect sensitive data. Victims are pressured into paying by the threat of publication. In this type of attack, backups alone do not remove the basis for extortion, as restoring data does not prevent stolen copies from being leaked. Defence priorities include securing internet-facing services, timely patching, and detecting anomalous data transfers.

FAMOUSSPARROW

DESCRIPTION: FamousSparrow is a China-aligned cyber espionage group assessed to have been active since at least 2019. Initially known for attacks on the hospitality sector, the group has also targeted government bodies, international organisations, engineering firms and law firms. Its objective is intelligence collection, and it uses custom malware developed for its operations.

RECENT ACTIVITIES: On 17 September, the Slovak cybersecurity company ESET published research on the group's espionage campaign targeting Latin America. Targets included government organisations in Argentina, Guatemala, Honduras, Panama and Peru, among others. The group had been using a new backdoor, SparroWocky, since at least August 2025, meaning that the activity was only reported in September but had been going on for considerably

longer. ESET assesses that the espionage is linked to China's interest in monitoring how states in the region respond to US pressure. One of the Panamanian targets was involved in a commercial dispute concerning ports in the Canal zone.

METHODS AND TACTICS: SparroWocky enables file theft, screenshot capture, command execution and collection of information about the infected machine. The malware is launched by abusing a legitimate program that loads a library modified by the attacker (DLL sideloading). The actual malicious payload is decrypted directly into memory, which hinders detection. Persistence can be achieved through a Windows service or an autorun registry key. Given these techniques, defenders should also monitor trusted programs for anomalous behaviour, as well as new services and unusual outbound connections.

PANZER RANSOMWARE

DESCRIPTION: Panzer is a new, financially motivated ransomware actor whose leak site was first observed on 5 August 2026. It offers malware and extortion tooling as a service to affiliates. According to its published business model, the affiliate carrying out the attack receives 80 per cent of the ransom, while the operators retain 20 per cent. Independently verified information on the group's background and technical capabilities remains limited.

RECENT ACTIVITY: According to a CyberXTron report published in early September, Panzer had named 16 victims from 11 countries on its leak site. Technology and manufacturing companies featured prominently in the listings. In August, the group named two Italian victims: the kitchen manufacturer Doimo Cucine and the telecommunications

engineering firm NTE Italia. Neither company had publicly confirmed an incident, so the information is based solely on the group's own claims.

METHODS AND TACTICS: Panzer promotes a model that combines data theft with file encryption and pressure applied via its leak site. Affiliates are offered an administration panel, a negotiation channel and payment-processing tools. The group advertises support for Windows, Linux, VMware ESXi and FreeBSD environments, but no independently analysed malware samples were available when the source report was published. The initial access vector had not been confirmed either. Key preparedness measures include securing remote access, restricting administrative privileges, isolating backups and monitoring large or anomalous data transfers.

SOURCES:



1 EVENTS IN THE CYBER LANDSCAPE

Cyberwatch Finland weekly reviews in September 2026

<https://www.reuters.com/legal/government/us-companies-face-rise-cyber-attacks-2026-08-26/>

<https://euobserver.com/235865/swedish-election-not-a-priority-for-moscow-interference-but-still-a-target-for-troll-farms-and-disinformation/>

<https://www.euronews.com/2026/09/16/thousands-of-suspicious-accounts-followed-swedish-political-voices-researchers-say>

<https://www.securityweek.com/japan-dismantles-first-north-korean-laptop-farm-as-us-and-allies-detail-wider-scheme/>

<https://www.patrick-breyer.de/en/chat-control-2-0-trilogue-update-no-mass-scanning-deal-for-now-european-parliament-stands-firm-against-search-plans/>

<https://www.kyberturvallisuuskeskus.fi/fi/uutiset/kyberturvallisuuskeskuksen-viikkokatsaus-402026>

<https://www.reuters.com/world/china/china-us-agree-30-billion-tariff-cut-ai-dialogue-during-xi-visit-2026-09-26/>

<https://www.arctictoday.com/pro-russian-hackers-declare-cyberwar-on-norway/>

<https://edition.cnn.com/2026/09/01/europe/leipzig-drone-germany-russia-intl>

<https://www.bbc.com/news/articles/cwn8mj131p5yo>

<https://www.dw.com/en/fact-check-false-claims-disinformation-operation-overload-afd-saxony-anhalt/a-79093254>

2 IN THE SPOTLIGHT

2.1 Vulnerability Numbers Continue to Rise

<https://www.securityweek.com/google-ai-is-changing-the-pace-and-profile-of-vulnerability-discovery/>

<https://therecord.media/microsoft-patch-tuesday-september-2026>

<https://msrc.microsoft.com/update-guide/releaseNote/2026-Sep>

2.2 Fake Online Shops Are Becoming Increasingly Sophisticated and Widespread

https://investigations.nebty-id.com/doppelcart?verdict=fleet_confirmed

<https://nebty-id.com/en/doppelcart-fake-shop-network/>

<https://srlabs.de/blog/bogusbazaar>

<https://www.malwarebytes.com/blog/scams/2026/03/inside-a-network-of-20000-fake-shops>

<https://www.humansecurity.com/learn/blog/phish-n-ships-responsible-disclosure-and-healthy-skepticism/>

<https://www.bleepingcomputer.com/news/security/doppelcart-fraud-network-uses-119-000-fake-shops-to-steal-credit-cards/>

3 FOLLOW THESE

3.1 Is Post-Quantum Cryptography More Urgent Than Previously Thought?

<https://www.tivi.fi/uutiset/a/d125e284-5c30-4267-b5a8-40f83d090cf4>

<https://thequantuminsider.com/2026/08/24/japans-full-stack-neutral-atom-quantum-computer-is-operational/>

<https://www.caltech.edu/about/news/caltech-team-finds-useful-quantum-computers-could-be-built-with-as-few-as-10000-qubits>

3.2 Cyber Threats Loom Over the Upcoming Us Midterm Elections

<https://www.swktech.com/swk-cybersecurity-news-recap-september-2026/>

<https://www.dhs.gov/news/2026/09/24/dhs-announces-release-2026-election-infrastructure-security-plan>

<https://www.cisa.gov/sites/default/files/2026-09/2026-CISA-Election-Security-Plan-FINAL-508c.pdf>

<https://socradar.io/blog/2026-us-midterms-election-cyber/>

<https://transparency.meta.com/sr/H2-2026-adversarial-threat-report/>

<https://www.nytimes.com/2026/09/18/us/politics/russia-election-disinformation-us-intelligence.html>

4 SEPTEMBER 2026: AI BROKE LOOSE, LIMITS WERE SOUGHT AND AT TIMES IT ALSO PROVIDED PROTECTION

<https://openai.com/index/safety-overview-gpt-6-astra/>

https://en.wikipedia.org/wiki/OpenAI%E2%80%93HuggingFace_incident

<https://www.abc.net.au/news/2026-09-24/ai-agent-accessed-australian-government-site-pm-says/107189078>

<https://www.atlanticcouncil.org/dispatches/what-the-proposed-ai-slowdown-means-for-the-us-china-and-humanity-at-large/>

<https://www.pbs.org/newshour/world/china-and-u-s-agree-to-establish-ai-safety-channel-and-continue-trade-and-military-talks>

<https://techxplore.com/news/2026-09-anthropic-blocked-misuse-ai-biological.html>

<https://www.iranintl.com/en/202609131676>

<https://www.nokia.com/newsroom/espanix-deploys-nokia-deepfield-defender-to-protect-spains-internet-traffic-from-ddos-attacks/>

<https://www.helpnetsecurity.com/2026/09/24/apple-ios-27-impersonation-risk-detection/>

THREAT INTELLIGENCE

Cyberattack on Dustin Group:

<https://www.dustingroup.com/information>

<https://www.dustingroup.com/en/media/#/pressreleases/dustin-confirms-successful-opening-of-websites-and-customer-portals-3466549>

Revolut data leak:

<https://www.khaleejtimes.com/business/tech/revolut-confirms-sensitive-customer-data-breach>

Breach of a Japanese Digital Agency system:

<https://www.digital.go.jp/news/2026-0911-01>

<https://www.securityweek.com/240000-hit-by-data-breach-at-japans-digital-agency/>

THREAT ACTORS

Cl0p:

<https://www.cyber.gc.ca/en/guidance/profile-ta505-cl0p-ransomware>

<https://www.reuters.com/legal/government/cybercrime-feud-erupts-dark-web-notorious-group-claims-hijack-rivals-website-2026-09-21>

FamousSparrow:

<https://www.welivesecurity.com/en/eset-research/beware-sparrowock-backdoor-bites-commands-catch>

<https://therecord.media/china-hackers-latin-america-espionage>

Panzer Ransomware:

<https://cyberxtron.com/resources/blogs/panzer-ransomware-profile-of-an-emerging-double-extortion-operator-8102>

<https://andreafortuna.org/2026/09/07/panzer-ransomware-italian-victims/>



Cyberwatch MONTHLY

PUBLISHER Cyberwatch Finland
Kansakoulukatu 1b, 00100 Helsinki | www.cyberwatchfinland.fi

Netum & Cyberwatch CYBER RESILIENCE

We strengthen operational cyber resilience holistically by **IDENTIFYING** the most significant cyber threats, **ASSESSING** their business impact, and **ENSURING** that critical digital services and the operations that support them can withstand disruption.

➤ **Shared cyber and hybrid risk situational picture**

See threats. Assess risks. Secure services.



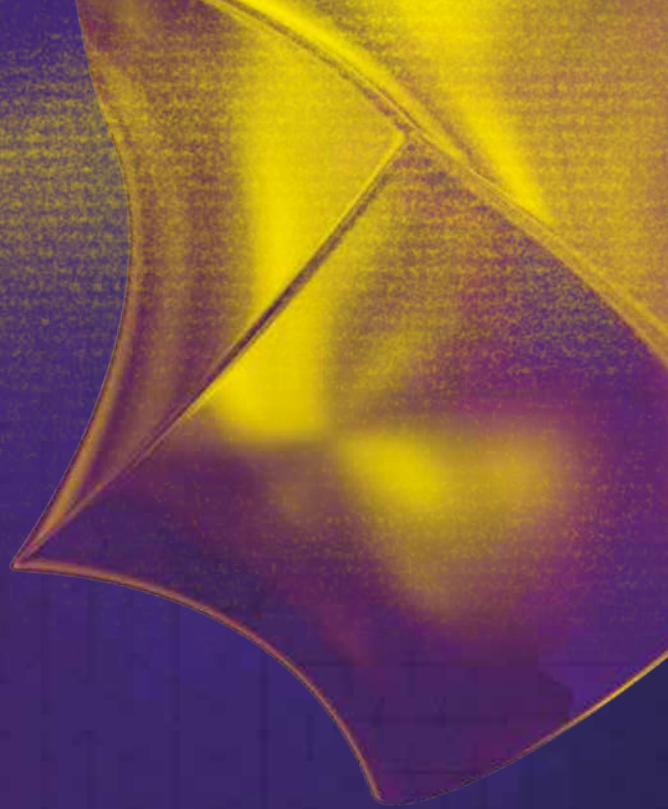
WE STRENGTHEN organisational and operational **CYBER RESILIENCE** – not just the technology management layer.

- ✓ Executive Situational Awareness Service
- ✓ Cyber Risk Management Services
- ✓ Intelligent Cyber Security & Data Protection – Consulting & Training services
- ✓ Digital Systems Security Services

Our offering is an evolving process that responds to the changing needs of the cyber landscape.

We make regulatory compliance easy for you.

Resilient business & crisis preparedness
Cybersecurity, defence & the European security landscape
Trusted digital society & democratic security



CYBER SECURITY NORDIC

25–26 Nov 2027
Helsinki Expo and Convention Centre

/MESSUKESKUS



Leading nordic meeting place for cybersecurity leaders, experts and decision-makers. And this is only the beginning. More than 80 leading cybersecurity and information security experts on stage. Explore the full event at cybersecuritynordic.com

