Penetration testing: are you testing what attackers target, and fixing what testers find?

Before the next test, write down the question it must answer and scope it around identity, internet-facing systems and critical apps, not an IP list. Judge each test by how many findings were fixed and retested on time.

Reviewed	Next review	Version
5 Nov 2026	**5 Dec 2026**	**1.0**

Written for

IT, security and engineering leads in EU organizations, many in Germany, who commission penetration tests or red team exercises, or must under DORA or PCI DSS

Latest version

sheets.einblick.dev/penetration-testing/

If you only do five things

1. **Start from a question, not a product.** The NCSC treats a penetration test as an audit of your vulnerability management: you should already know most of what it finds[1]. Write down the question the test must answer.
2. **Scope by attacker targets.** Exploits opened 32% of Mandiant's 2025 intrusions and voice phishing 11%, with groups calling help desks to get past MFA[2]. Scope identity, the help desk, the edge and critical apps first.
3. **Give testers the inside view.** The BSI recommends white-box tests, since black-box tests miss more and risk more damage[3]. Hand over network plans, test accounts, source code and your latest vulnerability assessment[3,1].
4. **Get written authorization from every owner.** The BSI says never to test without a written contract, and to include any hoster[3]. Microsoft won't authorize testers for you[4].
5. **Retest and track every finding to closure.** PCI DSS requires a retest to verify fixes[5] and DORA a check that every weakness is addressed[6]. Give each finding an owner and a deadline, and book the retest with the test.

Commission a test in this order

1. **Agree the scope with the risk owner.** Involve risk owners, technical staff and the testers, and agree boundaries, test types, effort in days and reporting[1].
2. **Get everyone on board.** The system owner, any hoster or SaaS provider and, if staff or personal data are affected, the data protection officer and works council[3].
3. **Set the rules of engagement.** Agree the window, exclusions, depth and a stop rule[3], and a contact on call[1]. Freeze changes; if a flaw can't wait, pause the test[3].
4. **Read the report critically.** The risk decision is yours[1]. Group repeated findings[3] and ask why you didn't already know about each one[1].
5. **Close the loop.** Fix by deadline, retest and add each finding to your own regular checks[3]. Removing the software or adding monitoring may suffice instead of the suggested patch[1].

Pick the test that answers your question

- **Vulnerability scan: which known flaws are open?** Automated and frequent; see the attack surface sheet. A penetration test adds a competent manual attacker[5].
- **Penetration test: does this system hold up against a skilled tester?** Days of manual work on one system with full information, after go-live and significant changes[3].
- **Assumed breach: how far does one compromised laptop or account get?** The tester starts inside, as in the NCSC's lost laptop scenario[1].
- **Red team: would you detect and stop a real attacker?** A covert, threat-led attack on live systems; under DORA the active phase lasts at least 12 weeks[7]. Wait until logging and alert handling work.
- **Purple teaming: which attacker techniques do you detect?** Testers and defenders work together and tune detections as they go. DORA requires it after every threat-led test[7].
- **Bug bounty: what do outsiders find over time?** See the secure development sheet.

Scope by risk, not by IP list

- **Identity.** Paths from a normal user to domain admin or Global Administrator, including AD CS; the privileged access sheet lists fixes.
- **The help desk.** A call asking for an MFA reset, under rules agreed with the works council and run by specialists[3].
- **The internet edge and critical apps.** Everything in the attack surface sheet's inventory, and apps tested as the API security sheet describes.
- **Backups, hypervisors and tenants.** Ransomware groups target backups and virtualization management[2]; Microsoft encourages tenant tests[4].
- **How often.** Test high-risk areas yearly and after major changes, as PCI DSS requires for card data[5]. For the rest, the BSI suggests every two to three years[3].

Choose a provider

- **Treat certificates as signals.** PCI DSS calls them an indication of skill and asks about past engagements too[5]. BSI personnel certification and CREST include practical exams[3].
- **In Germany, check the BSI list.** It names 22 providers certified for IS penetration tests[9].
- **In UK government and CNI, use CHECK.** Its team leaders hold the UK Cyber Security Council's Principal title and SC clearance[10].
- **Buy the people, not the logo.** Get CVs of the named testers, at least two, check they do the work and rotate them over the years[3].

Stay legal and safe

- **Authorization.** In Germany, getting past access protection without permission carries up to three years in prison or a fine[11].
- **Cloud.** Azure and Google Cloud need no notice to test your own resources, AWS none for its listed services[4,13,12]. Azure and AWS ban DoS; AWS must approve covert red teaming and phishing, asked two weeks ahead[12].
- **Production and OT.** No test rules out side effects[1]: keep a stop rule. In OT, use testers with OT experience[8] and test in outages or on a replica.

What changed in 2026

- **AI-assisted testing needs ground rules.** The NCSC's 10 questions from May 2026 cover who fixes the findings, sandboxing, data terms and legality[17]. Ask testers who use AI the same.
- **CREST accredits AI security testing.** It named the first four companies in November 2026[18]. Ask for it, or comparable evidence, before testing a GenAI or LLM system.

Tests a regulation or contract requires

NIS2 requires essential and important entities to assess whether their measures work[14], and cloud, managed service and other digital providers to set a risk-based testing policy[15]. DORA requires yearly tests of systems behind critical or important functions, except at microenterprises[6].

DORA threat-led penetration test (TLPT)

Entities their authority selects, at least every 3 years, on live systems[6]. In Germany the Bundesbank's TIBER team manages tests BaFin or the ECB orders[16]. Clocks start at the notice, then at the end of red teaming[7].

3 months	6 months	12+ weeks	4 weeks	10 weeks
Initiation papers	Scope document	Active red teaming	Red team report	Replay, purple teaming

PCI DSS v4.0.1, requirement 11.4

The card data environment, by a qualified, organizationally independent tester who needn't be a QSA[5]. For segmentation tests, see the network segmentation sheet.

12 months	After changes	After fixes
Internal and external tests	Significant upgrades	Retest exploitable findings

What you can skip

- **Black-box tests for realism.** Testers then spend paid days finding out what you could tell them. The BSI recommends white-box tests[3].
- **A red team before the basics.** If scans still find unpatched edge systems and nobody handles alerts, a red team only confirms it, at a high price.
- **Paying testers to recheck every finding.** The BSI wants your own staff to check fixes with repeatable tests[3]. Book external retests for critical and high findings only.

Mistakes that cost the most

- **Taking a clean report as proof.** PCI DSS says a test that finds nothing usually reflects the tester's shortcomings[5]; see the edge devices sheet.
- **Letting the WAF test itself.** If it blocks the testers, you test the filter. Lift the blocks for them and test the gateway separately[3].
- **Leaving the testers' tools behind.** DORA makes testers remove backdoors and delete captured credentials[7]; demand the same.
- **Findings without an owner.** Risk decisions stay with you, not the testers[1].

What this looks like at your size

Small team, no security staff

- Test your cloud tenant and internet edge yearly.
- Fix critical and high flaws in 30 days, 14 if exposed.
- Shortlist from the BSI list or CREST.

Growing, with IT or security staff

- Test AD and Entra ID yearly from a user's laptop.
- Test critical app releases with source code.
- Report open findings by age every month.

Enterprise or regulated

- Plan tests by risk for several years ahead.
- Map the plan to DORA[6] and PCI DSS 11.4[5].
- Report time to close to the board.

Version history

5 Nov 2026, version 1.0	Published after a copy edit and a source-by-source fact check, which aligned fix deadlines with the patching sheet and narrowed 12 statements to what the sources say.

Sources

1. Penetration testing. NCSC (UK), reviewed 10 Jan 2022. ncsc.gov.uk
2. M-Trends 2026. Google Cloud (Mandiant), 23 Mar 2026. cloud.google.com
3. Praxis-Leitfaden für IS-Penetrationstests, version 1.2. BSI, Nov 2016. bsi.bund.de
4. Azure penetration testing. Microsoft Learn, 15 May 2026. learn.microsoft.com
5. PCI DSS v4.0.1, req. 11.4. PCI SSC, Jun 2024. pcisecuritystandards.org
6. Regulation (EU) 2022/2554 (DORA), Art. 24 to 27. EUR-Lex, 2022. eur-lex.europa.eu
7. Delegated Regulation (EU) 2025/1190 (DORA TLPT). EUR-Lex, 18 Jun 2025. eur-lex.europa.eu
8. Building more resilient CNI: what industry pen testers told us. NCSC (UK), 1 Jul 2026. ncsc.gov.uk
9. Liste IS-Penetrationstests. BSI, checked 5 Nov 2026. bsi.bund.de
10. Information for CHECK buyers. NCSC (UK), checked 5 Nov 2026. ncsc.gov.uk
11. StGB § 202a. BMJV, checked 5 Nov 2026. gesetze-im-internet.de
12. Penetration Testing. AWS, checked 5 Nov 2026. aws.amazon.com
13. Cloud Security FAQ. Google Cloud, checked 5 Nov 2026. support.google.com
14. Directive (EU) 2022/2555 (NIS2), Art. 21. EUR-Lex, 2022. eur-lex.europa.eu
15. Implementing Regulation (EU) 2024/2690, Annex 6.5. EUR-Lex, Oct 2024. eur-lex.europa.eu
16. Implementation of TIBER-DE, version 4.0. Bundesbank, Jul 2025. bundesbank.de
17. 10 questions to ask when using AI models to find vulnerabilities. NCSC (UK), 11 May 2026. ncsc.gov.uk
18. First companies accredited to test the security of AI systems. CREST, 2 Nov 2026. crest-approved.org