How to Plan Your Penetration Testing Program

A strategic framework for security leaders: prioritize testing, control costs, and build a security posture that's constantly improving.

Penetration Testing Is Not an Event

You cannot fix what you cannot measure. Penetration testing is that measurement. It tells you, in concrete terms, where your defenses hold and where they fail. But a single measurement gives you a snapshot—and your environment is not static. Cloud configurations drift. Code changes daily. Systems are constantly being patched often changing configuration files. Identity boundaries expand through SaaS integrations and third-party access. Privilege creep accumulates silently. Attack techniques adapt continuously. A measurement taken six months ago may already be stale.

Because the environment keeps changing, the measurement must be ongoing. Organizations that achieve strong security postures treat penetration testing as a programmatic, iterative discipline—not an annual checkbox. Each round of testing builds on the last, revealing structural weaknesses that isolated engagements never surface. Over multiple cycles, you begin to see risk patterns: recurring privilege escalation paths, persistent architectural fragility, process gaps that resist one-time fixes. These patterns are where the real value lies, and they only emerge through sustained, strategic testing.

Think in Three Dimensions: Surface × Level × Time

Every penetration testing decision sits at the intersection of three dimensions. If any one of them is missing from your planning, you will either overspend, leave critical gaps, or both.

- **Surface** — the technology landscape exposed to risk. Not just an inventory of assets, but an exposure model: what is reachable, what amplifies privilege, and what causes the greatest damage if compromised.
- **Level** — the depth of adversarial rigor applied. Not all surfaces need the same intensity. The level should match the business value and threat exposure of each surface, not follow a one-size-fits-all escalation.
- **Time** — the maturity path and cadence across years. Testing should be phased and progressive, with formal reevaluation at regular intervals to account for business changes, emerging threats, and evolving architecture.

Every testing decision is a coordinate in this model. When you plan in all three dimensions simultaneously, testing becomes structured rather than reactive, portfolio-based rather than siloed, and progressive rather than episodic. This is how you eliminate blind spots, avoid misallocated budgets, and prevent the false assurance that comes from testing the wrong things at the wrong depth.

Surface: An Exposure Model, Not an Asset List

Listing your technology categories is a starting point, but the real question is how these surfaces connect to create exploitable paths. An attacker does not care about your org chart or your asset inventory. They care about reachability, privilege amplification, and lateral movement.

One size does not fit all. Each surface in your environment carries a different risk profile and requires different testing capabilities and depth to evaluate properly. A web application assessment demands expertise in business logic and authentication flows. An ICS/SCADA evaluation requires operational technology specialists who understand safety-critical environments. Cloud testing needs analysts who think in IAM policies and lateral movement across services. The testing categories below reflect these distinctions—your job is to identify which exist in your environment and, critically, how they interconnect.

Example: A compromised credential (Identity) → VPN access (External Surface) → internal application with elevated permissions (Internal App) → database containing customer PII (Crown Jewel). Each hop crosses a different surface. If you test each surface in isolation, you will never see the chain.

Testing Surface	What It Covers	Typical Starting Level
Credential Testing	Password policies, credential hygiene, authentication mechanisms	Level 1 – baseline credential strength and exposure
External Attack Surface Testing	Internet-facing perimeter: DNS, mail, VPN, public IPs, firewalls	Level 1 – what attackers see first; fast, high-value start
Network & Data Center Penetration Testing	Internal LAN/WAN, segmentation, lateral movement, Active Directory	Level 1 (branch offices); Level 2 (HQ, data center)
Cloud Penetration Testing	AWS/Azure/GCP configs, IAM, storage, serverless, container, segmentation, lateral movement	Level 2 – mis-configs chain fast in cloud; needs human analysis
Web Application Penetration Testing	Customer portals, SaaS, e-commerce, internal apps with business logic	Level 2 – custom logic requires manual testing
API Penetration Testing	REST, GraphQL, SOAP APIs; auth, authorization, data exposure	Level 2 – APIs are increasingly the primary attack surface
IoT/Penetration Testing	Connected devices, sensors, cameras, building systems	Level 1 (fleet); Level 2 (critical/medical IoT)
ICS/OT Penetration Testing (SCADA)	Operational technology, industrial controls, SCADA networks	Level 2 minimum – safety-critical; specialized expertise
Physical Penetration Testing	Badge access, tailgating, lock bypass, social engineering at entry	Level 3 – entirely human-driven; tests people and processes
Tools/Services/Controls	SOC effectiveness, EDR/MDR validation, SIEM, DLP, deception tech	Level 3 – validates whether your security stack stops real attacks

Level: Depth Is Strategic, Not Automatic

Not every surface requires Level 3. Not every surface will progress beyond Level 1. Escalation depends on business criticality, threat exposure, and organizational cybersecurity maturity—not on a default assumption that deeper is always better. A remote branch office may remain Level 1 indefinitely. Your production database may start at Level 1 for a baseline and reach Level 3 within 12 months. Some environments like cloud and web applications should begin at Level 2 because their complexity demands it.

Level	What You Get	Best For	Escalation Criteria
Level 1	Rapid, broad-surface assessment. Analyst-guided automation. Surfaces high-probability risks and closes obvious gaps.	Baselines, branch offices, IoT fleets, routine hygiene, post-remediation checks	Sufficient for low-criticality assets. Escalate if findings suggest deeper chaining risk.
Level 2	Intent-driven pen testing. Human analysts chain vulnerabilities, simulate attacker behavior, map real exploit paths.	Data centers, cloud, web apps, APIs, ICS/SCADA, core networks	Required when business impact of compromise is high. Escalate to L3 only when specific triggers are met (see below).
Level 3	Full adversary emulation. Custom exploits, stealth tactics, specific objectives. Tests people, processes, and technology as one system.	Crown-jewel assets, SOC validation, physical security, control & security tool(s) effectiveness	Reserved for highest-value targets. Not all surfaces will ever reach this level.

When Does Level 3 Make Sense?

Level 3 is important when it is important—not as a default progression. Specific trigger events drive the decision:

Trigger	What It Means
Board or regulatory requirement	The board, a regulator, a key customer, or an insurance underwriter requires adversary-level validation of your crown-jewel defenses. This is a governance decision, not a technical one.
Diminishing returns at Level 2	Your Level 2 testing consistently returns few or no critical and high findings, and even medium findings are at low levels. Your program is mature. Level 3 is how you validate that your crown jewels are truly protected against a motivated, skilled adversary—not just against known vulnerability patterns.
Crown-jewel assurance	You need to be certain—not just confident—that the assets which would cause existential damage if breached (financial systems, IP, customer data at scale) can withstand a targeted, persistent attack across people, process, and technology.

Decision factors for level assignment: data sensitivity, exposure footprint (external vs. internal), regulatory burden, business continuity impact, and whether the surface connects to other high-value systems.

A Note on Remediation Realism

In an ideal world, critical findings are resolved within 10 days, highs within 30, and mediums within 90. That is the standard most frameworks recommend. That is not reality for most organizations, especially when internal resources are constrained and stretched across competing priorities. Remediation capacity—not testing capacity—is typically the bottleneck.

This is where prioritization becomes critical. Not all findings carry equal business risk, and attempting to fix everything simultaneously leads to nothing getting fixed well. Scapien's reporting is designed to support this: we prioritize findings by actual exploitability and business impact, so your team knows exactly where to focus limited resources for maximum risk reduction. Where internal capacity is insufficient, we help identify what can be deferred safely and where bringing in remediation partners makes sense.

Time: Plan in Three-Year Cycles, Reevaluate Every 18 Months

A 12-month testing plan is tactical. A three-year plan is strategic. It enables capital allocation discipline, prevents oscillation in priorities, aligns testing maturity with business growth, and supports trend analysis across testing cycles.

A word of realism: the timeline below assumes a typical starting posture. Organizations with mature security programs may compress it significantly—some will run adversary simulation on crown jewels in Year 1. That is achievable, but it requires a serious commitment to remediation between testing rounds. Be honest about what your team can execute, and plan to bring in partners for surge remediation if needed. Ambitious timelines without remediation follow-through produce expensive reports that sit on a shelf.

Phase	What to Do	Why
Year 1 Baseline & Surface Map	Level 1 across full external surface, credentials, and branch networks. Level 2 on business-critical cloud, web apps, and APIs. Identify the complete attack surface. Remediate high-probability findings.	Establishes your starting position. Eliminates the low-hanging fruit responsible for most breaches. Validates which surfaces warrant deeper testing in Year 2.
18-Month Reevaluation	Formal strategic review with your testing vendor. Evaluate progress against Year 1 objectives. Reassess surfaces (new SaaS, acquisitions, architecture changes). Adjust levels and priorities for Year 2–3. Reprioritize based on emerging threats and business shifts.	Fixed plans decay. Business changes, threat landscapes shift, and new surfaces emerge. The 18-month checkpoint prevents static models, legacy priorities, and drift between your testing program and your actual risk.
Year 2 Escalation on Critical Surfaces	Elevate to Level 2 on data centers, core internal networks, and ICS/SCADA if applicable. Begin Level 3 scoping on crown-jewel assets. Re-test all remediated findings from Year 1.	Deepens rigor where business exposure is highest. Validates that Year 1 remediation actually closed the gaps. Begins to reveal structural risk patterns across cycles.
Year 3 Adversary Simulation	Level 3 on crown jewels: production databases, financial systems, IP repositories. Physical pen testing. SOC and control validation. Maintain Level 1–2 recurring cadence on other surfaces.	Tests your highest-value targets the way a real adversary would. Multi-cycle data now reveals systemic patterns, recurring exploit classes, and architectural weaknesses that isolated tests never surface.

After Year 3, restart the cycle. The strategic plan should be formally reviewed every 18 months in partnership with your testing vendor, evaluating progress against stated objectives, incorporating architecture and business changes, and adjusting surfaces and levels accordingly. Surfaces may shift. Levels may escalate or de-escalate. The model adapts; the discipline does not.

Three Dimensions in Practice

The following examples show how Surface × Level × Time coordinates work in real planning decisions.

Scenario	Surface	Level Path	Time	Reasoning
Regional branch offices (10 locations)	Network, External	Level 1 → Level 1	Yr 1 → Ongoing	Standardized environments, limited data, low business-continuity impact. Level 1 annual cadence is sufficient. No escalation needed unless findings change the picture.
Customer-facing SaaS platform	Web App, API, Cloud	Level 2 → Level 2 → Level 3	Yr 1 → Yr 2 → Yr 3	High exposure, complex logic, customer data at risk. Starts at Level 2 because automated scanning alone cannot evaluate custom business logic. Escalates to Level 3 once baseline remediation is complete.
Production database cluster (financial records)	Network, Cloud, Controls	Level 1 → Level 2 → Level 3	Yr 1 → Yr 1–2 → Yr 2–3	Crown jewel. Rapid L1 baseline in Q1, L2 within 6 months to find exploit chains, L3 adversary emulation by Year 2–3. Cross-surface testing matters here: credential → VPN → lateral → database.

Penetration Testing as Strategic Feedback

Done right, penetration testing is not just a security exercise. It is a decision-support tool that provides four types of strategic intelligence:

Exploitability validation	Confirms which theoretical risks are actually exploitable in your environment—separating real threats from noise.
Architectural feedback	Reveals structural weaknesses in how systems connect, segment, and authenticate—issues that patching alone will not solve.
Budget prioritization	Shows where your next security dollar will have the greatest impact, based on demonstrated exploit paths rather than theoretical scores.
Control effectiveness	Measures whether your security stack, processes, and people actually detect and stop attacks—not just whether they are deployed.

When you treat testing output this way, it drives remediation investment, identifies systemic weaknesses, and enables architectural strengthening. It stops being a compliance artifact and starts informing strategic security decision-making.