ALACRINET / OFFENSIVE INTELLIGENCE UNIT

Pen Test Vendor Evaluation Checklist

12 questions most pentest vendors can't answer.

Real attack paths, not scanner output.

Named operators, not offshore teams.

Honest answers, or a reason to walk.

Bailey Besheer
Managing Director of Cybersecurity Services, Alacrinet
pentesting.alacrinet.com

What to Watch For on the Call

The questions are written so real pentest shops can answer them specifically and confidently. Commodity vendors will dodge the question, deflect to a sales engineer, or give you something vague. Two things to watch for in every answer:

1. Can they answer without a sales engineer?

A serious shop has these answers ready in real time. If they need to check with someone, that tells you something.

2. Is the answer specific or generic?

Real shops give you numbers, named tools, named frameworks, and concrete examples. Commodity shops give you capability statements and "it depends."

We don't follow up unless you ask. This checklist is the only thing we'll send you. No drip campaign. No "checking in" emails. No sales rep assigned to your account. If you want to talk to us, contact details are at the back. If you don't, you'll never hear from us.

The 12 questions are organized into four sections: **Methodology, Team and Credentials, Deliverable Quality,** and **Engagement Terms.**

SECTION 1: METHODOLOGY

Question 1 of 12

What percentage of findings are manually validated before reporting? What percentage of exploitation is performed by hand?

Why it matters. These are two different metrics that commodity vendors love to blur together. Manual validation is the bar for findings: every CVE, every misconfiguration, every credential in the report should have been confirmed by a human before it shipped. Manual exploitation is the bar for testing: how much of the actual attack-chain work was done by an operator instead of an automated scanner.

STRONG ANSWER

100 percent of findings manually validated before reporting. Manual exploitation typically 60 to 90 percent depending on engagement scope, with some discovery and reconnaissance appropriately tooled. The vendor can articulate exactly what gets tooling, what gets validated by hand, and why.

RED FLAG

Findings list contains thousands of unvalidated CVEs. Vendor can't distinguish between scanner-discovered, scanner-validated, and manually-exploited findings. "We use a combination of approaches." "It depends on the engagement."

SECTION 1: METHODOLOGY

Question 2 of 12

Which methodology framework do you follow? Can you map findings to MITRE ATT&CK?

Why it matters. PTES, NIST 800-115, and OSSTMM are the recognized methodology standards in this space. Mapping findings to MITRE ATT&CK is the minimum bar for a credible deliverable. A shop that can't name a framework is improvising.

STRONG ANSWER

Names a specific framework (typically PTES) and can describe how findings are mapped to ATT&CK techniques.

RED FLAG

"We follow industry best practices." "Our methodology is proprietary." No working familiarity with ATT&CK.

SECTION 1: METHODOLOGY

Question 3 of 12

Do you run vulnerability scanners during the engagement, and is scanner output included in findings?

Why it matters. Real pentest shops use scanners selectively for reconnaissance and validate every finding manually. Commodity shops include raw scanner output in deliverables, which inflates finding counts with low-quality false positives.

STRONG ANSWER

A specific position on scanner use. Real shops do not deliver raw scanner output as findings. They may use Nmap for port discovery and exclude vulnerability scanners (Nessus, Qualys, Rapid7) from final reporting.

RED FLAG

Findings list contains thousands of CVEs from scanner output. Unable to articulate the line between scanner findings and manually validated findings.

SECTION 2: TEAM AND CREDENTIALS

Question 4 of 12

Where is the testing team based, and what is the seniority distribution on this engagement?

Why it matters. Commodity engagements are often staffed by offshore or junior testers with one senior reviewer. A US-based, all-senior team carries different cost and meaningfully different output quality.

STRONG ANSWER

Specific location of testers (US-based, EU-based), specific seniority distribution, and willingness to disclose who will actually be on the engagement before contract signing.

RED FLAG

"We have a global delivery team." Unwilling to specify locations. "We'll assign appropriate resources."

SECTION 2: TEAM AND CREDENTIALS

Question 5 of 12

What certifications do your testers hold?

Why it matters. OSCP, OSEP, and OSCE3 are hands-on testing certifications from Offensive Security. CRTO and CRTL are red-team certifications from Zero-Point Security. IAT III is the DoD operator-level cyber certification, requiring hands-on practical demonstration. CSIS is a hands-on operator credential. These are the credentials that map to actual pentest work. CISSP and Security+ alone are governance and management certs. They're useful, but they're not pentest credentials by themselves. The team's cert mix tells you whether you're hiring operators or auditors.

STRONG ANSWER

Names specific hands-on operator certifications (OffSec, SANS GIAC hands-on, Zero-Point, DoD IAT III, CSIS), with depth across the team.

RED FLAG

Only governance certs (CISSP, CISM, Security+) with no hands-on credentials behind them. Vague references to "industry certifications." Certifications listed are sales-focused or management-oriented.

SECTION 3: DELIVERABLE QUALITY

Question 6 of 12

Can you provide a sample report template showing how findings are presented, attack paths chained, and remediation guidance written?

Why it matters. This question separates real shops from commodity shops more cleanly than anything else on the list. A real pentest report template demonstrates attack-narrative structure, exploitation-evidence formatting, chained attack-path documentation, and prioritized remediation guidance. A commodity template is a CVSS-sorted list of CVEs with vendor patch advisories pasted in. You do not need to see a past client's data to evaluate the structure. A vendor handing you a redacted past client deliverable is also telling you something about their discretion practices, which is what Question 12 covers.

STRONG ANSWER

Yes. Vendor provides a synthetic or sanitized template that demonstrates attack narratives, chained paths, exploitation-evidence formatting, and remediation guidance structure. No client data involved.

RED FLAG

"We can't share samples for confidentiality reasons," with no alternative offered. Sample is a generic vulnerability list with no narrative or exploitation evidence. Or: vendor hands over past client work with names redacted, which signals weak discretion practices on real engagements.

SECTION 3: DELIVERABLE QUALITY

Question 7 of 12

How does the report distinguish demonstrated attack paths from theoretical vulnerabilities?

Why it matters. A vulnerability is a possibility. A demonstrated attack path is a proven impact. Real pentests prioritize the latter; commodity shops can only deliver the former.

STRONG ANSWER

Discusses how the report ties individual findings into chained attack paths, with explicit notation of which findings were exploited end-to-end versus identified but not exploited.

RED FLAG

All findings presented as standalone vulnerabilities. No narrative tying findings together. Unable to articulate the difference between a vulnerability and an attack path.

SECTION 3: DELIVERABLE QUALITY

Question 8 of 12

What's the typical finding count for an engagement of our size, and what percentage are critical or high?

Why it matters. Scanner-based deliverables produce thousands of findings, most of them informational or low-severity. Real pentest deliverables produce dozens of curated findings, with a much higher proportion of critical and high entries because each one has been validated.

STRONG ANSWER

Specific numbers. For an enterprise engagement: typically 40 to 150 findings, with 10 to 30 percent rated critical or high. The vendor should be able to give a ballpark without consulting anyone.

RED FLAG

"We find as many as are present." Very high finding counts (thousands). Unable to give a rough range without checking delivery.

SECTION 3: DELIVERABLE QUALITY

Question 9 of 12

Do findings include reproduction steps, screenshots, and commands?

Why it matters. This separates work that helps your team remediate from work that just documents the problem.

STRONG ANSWER

Each finding includes proof-of-concept content, reproduction commands, screenshots showing exploitation, and remediation guidance specific to your environment.

RED FLAG

Findings only reference CVE IDs and CVSS scores. Remediation advice is generic ("apply vendor patches"). No proof of exploitation is included.

SECTION 4: ENGAGEMENT TERMS

Question 10 of 12

What is your retest policy? Is there a time cap on remediation validation?

Why it matters. A pentest is only valuable if findings get fixed. Most shops cap remediation testing at 30 or 90 days. That's almost never enough time for an internal team to actually work through a large finding list. The retest policy tells you whether the vendor is in the relationship for the deliverable or for the outcome.

STRONG ANSWER

Unlimited or very long retest windows. No per-finding cap. Willingness to validate fixes as your team works through them. Past examples of staying engaged beyond standard windows without additional billing.

RED FLAG

Hard 30-day or 90-day cap. Each retest is a separate billable engagement. No proactive offer of validation support.

SECTION 4: ENGAGEMENT TERMS

Question 11 of 12

If a critical-severity issue surfaces mid-engagement, what's your escalation process?

Why it matters. Real pentest engagements regularly surface critical findings that can't wait for the final report. Credentials sitting in cleartext. Active exploitation paths into your environment. Evidence of prior compromise. Good vendors call you the same day. Bad ones bury it in the report a month later.

STRONG ANSWER

A defined notification SLA (24 hours is typical), named escalation contacts on both sides, and a real-time channel (Slack, Teams, or signed email) for events that can't wait.

RED FLAG

"We document it in the report." No SLA. No interim communication mechanism. Critical findings handled the same as informational findings.

SECTION 4: ENGAGEMENT TERMS

Question 12 of 12

What happens to client testing artifacts after the engagement ends?

Why it matters. Testing generates the most sensitive material your security program will ever produce: exploitation evidence, captured credentials, working notes, and raw data on what's exploitable in your environment. What the vendor does with that material after the engagement ends is a question of operational discretion, not just legal compliance.

STRONG ANSWER

A documented destruction policy with specific timeline (typical: 30 days post-engagement). All raw data, exploitation evidence, captured credentials, and working notes destroyed. No retention for case studies or marketing reuse.

RED FLAG

Vague answer. "We store it securely." No commitment to a destruction timeline. Vendor publishes case studies or anonymized engagement summaries. A redacted narrative is still a narrative. The wrong reader can recognize the environment.

About: How Alacrinet's Offensive Intelligence Unit Answers These Questions

> *"Alacrinet operates at the intersection of offensive security and discretion, and the discretion is non-negotiable. Clients engage us during moments of operational vulnerability: when something has happened, when something is about to, when leadership needs an honest read on exposure that internal teams can't deliver. They trust us with a level of access that almost nothing else in their security program receives. That trust is not material for marketing."*
>
> **Bailey Besheer,** Managing Director of Cybersecurity Services

Most pentest vendors can't answer most of these questions honestly. Alacrinet's Offensive Intelligence Unit is built so we can.

Senior, US-Based Operators

Our testers are senior, US-based, and do exploitation by hand.

PTES + MITRE ATT&CK

We work to PTES methodology and map findings to MITRE ATT&CK.

Attack-Path Reports

Reports lead with demonstrated attack paths rather than vulnerability lists.

Unlimited Retest

Remediation validation is unlimited and uncapped.

24-Hour Escalation

Critical findings get escalated inside 24 hours rather than buried in a final report.

30-Day Destruction

Testing artifacts get destroyed 30 days after engagement end, with no exceptions for case studies or marketing reuse.

If you're evaluating vendors and want a direct read on how Alacrinet's approach compares to what you're being offered, we'll do the comparison with you. No pitch, no nurture sequence. Just the technical assessment of where each proposal lands on the questions in this document.

ALACRINET / OFFENSIVE INTELLIGENCE UNIT

Get a Second Read on Your Vendor Proposals

Send us the proposal you're evaluating and we'll mark it up against the 12 questions in this checklist. You get a candid technical read in writing within 48 hours. We don't ask for a meeting and we don't follow up unless you ask us to.

Schedule a Call

SCHEDULE A CALL

calendly.com/bailey-besheer-alacrinet/30min

EMAIL

hello@alacrinet.com

WEB

pentesting.alacrinet.com

Built by Alacrinet's Offensive Intelligence Unit. The questions in this document are the same operational standards we hold ourselves to. We're publishing them because the gap between real penetration testing and commodity vulnerability scanning keeps widening, and buyers need a way to tell the difference. Use it on any vendor. Including us.

More on our discretion policy: **pentesting.alacrinet.com/why-no-case-studies**

Bailey Besheer LinkedIn
Managing Director of Cybersecurity Services, Alacrinet