aperiops

Cybersecurity for AI Builders

Cybersecurity is a business enabler. When done right, it supports business continuity, resilience, and redundancy. It protects your business, your reputation and your customers.

1. The Three Strategic Goals

Instead of just chasing every new attack vector, focus on the system's foundations and apply security concepts in a simple way suited for your development stage.

- **Prevention**: Focus on *Attack Surface Minimization (ASM)*. Make it as difficult as possible for attackers to penetrate the network by securing all resources exposed to the world (the outer defense ring).

Operate under an 'Assumed Breach' mindset.

- **Containment**: If an attacker gets in, how do you stop lateral movement or privilege escalation? Use tools to identify and contain the threat immediately.

- **Recovery**: If defenses fail, the goal is to return to normal business activity as fast as possible. This requires remote backups and an Incident Response Procedure (IRP).

2. The Layers Approach: Code Doesn't Live in a Vacuum

AI has increased development speed, but at a cost: code created by AI is often overly complex and lacks secure development best practices.
But your application isn't just code - it relies on existing infrastructure and third-party services.

- **Public Facing (Outer Layer):** Focus on Rate Limiting, Throttling, MFA, and Secrets management. Use strict CORS policies and validate request origins.

- **Interface**: Sanitize all input fields. For AI, add content filtering and multimodal cleansing (e.g., removing hidden commands from files).

- **Application Logic:** Separate system prompts from user inputs. Implement 'Human-in-the-Loop' (HITL) for sensitive actions and collect logs of all agent activities for anomaly detection.

- **Core Layer**: Apply Least Privilege (Tool Scoping). Use parameterized operations; never run raw strings directly against the database. Limit the LLM's access to core layer elements to the absolute minimum. Implement limited 'read-only' access and highly restricted 'modify' or delete permissions.

3. The Implementation Process

Security is a marathon, not a sprint. Follow these steps to build your security posture:

1. **Asset Management**: Map what you have and where it is (databases, files, versions, licenses, certificates). Establish a Single Source of Truth.
   *The 'Single Source of Truth' for your assets must be **extra-secured** and serves as the foundation for your company's growth.*

2. **User Management**: Define how *users* and *admins* are created, how permissions are granted, and how they are deleted. Ensure you're not violating compliance requirements.
   Automation is easier once the process is defined.

3. **Apply CIA Model**: Evaluate every interface and system based on Confidentiality, Integrity, and Accessibility.

4. **Monitoring & Alerts**: Activate logs and set up alerts.
   Don't fear the alerts- they are usually easier to understand and fix than they seem.

5. **Choose a Path**: Consult with an AI assistant to learn basic defense frameworks, then choose the one that fits your specific business logic and systems. This can change over time.

6. **Reinforce Infrastructure**: One tool at a time - improve configurations, check the built-in security features of the tools you use or complementary solutions from the same vendors. Look for '*low hanging fruit*'.

Key Strategic Highlights

- If you are required to or choose to be aligned with a security standard or framework, add it to your AI assistant's context. If you have compliance requirements, it's recommended to work with an advisor.

- Don't give your AI access to your actual passwords or secrets, just give context on where they exist. Never store secrets or API keys in `.env` files uploaded to Git.

- Implement Rate Limiting/Throttling to prevent DoS attacks and high API costs.

- Security is a managed business process - track findings and set reminders to check systems in the future.

Incident Response: What to do if Compromised

- **DO NOT panic**.
- DO NOT delete evidence.
- DO NOT post about it publicly.
- Use your IRP.
- Call a professional.

No one knows your company better than you. You have full ownership. Challenge every recommendation your AI assistant is giving you.

For additional questions and follow-ups: advaha@aperiops.com | LinkedIn | X (Twitter)