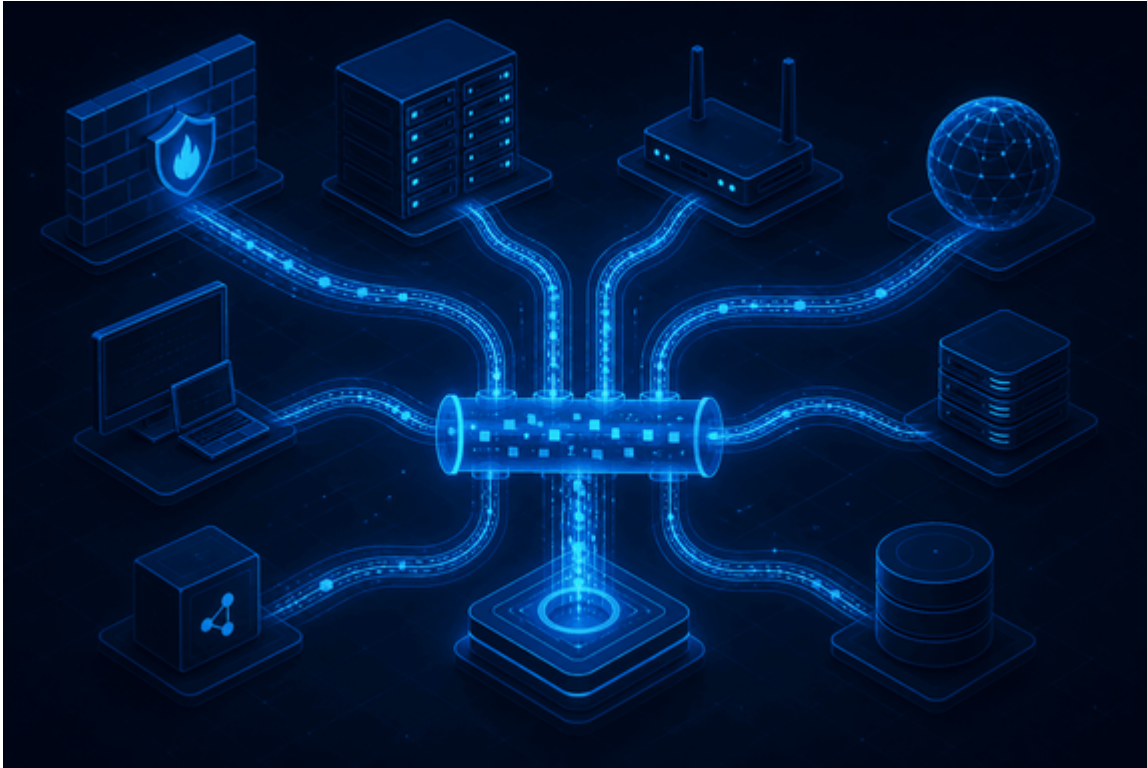


Log Sources: What to Collect Before You Can Analyze Anything

Sat, Aug 15, 6:00 PM EDT · <https://scottslab.io/posts/log-sources-and-data-types>



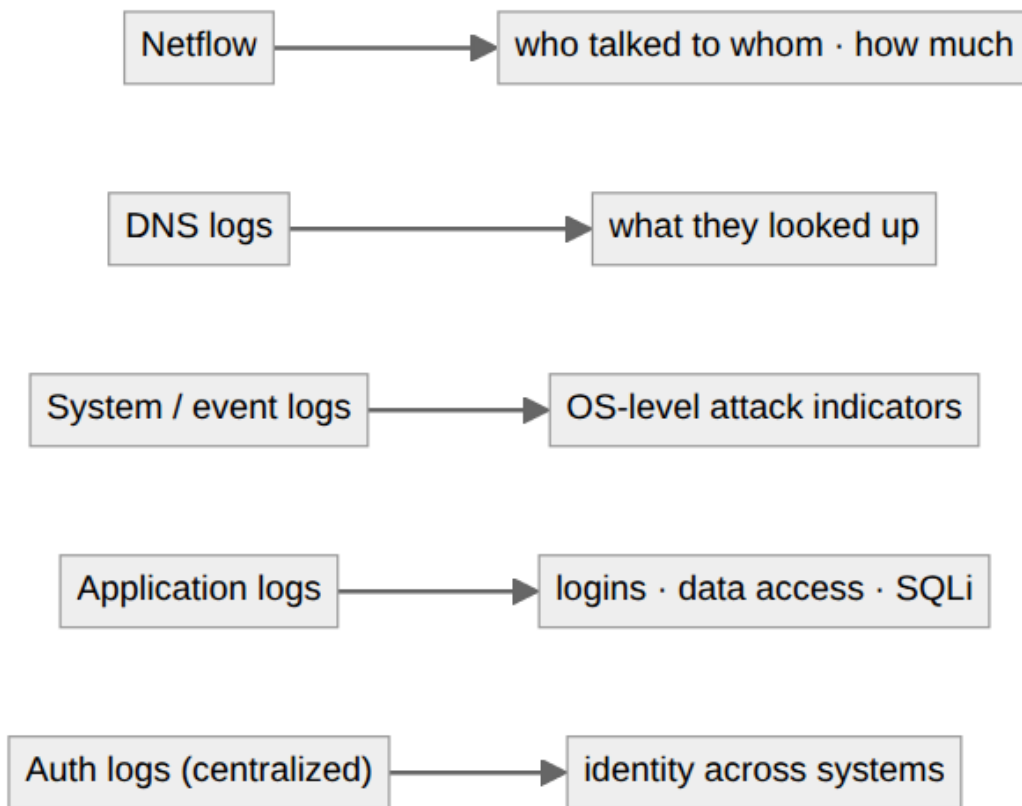
TL;DR

You can only investigate what you logged, and each source answers a different question: netflow tells you who talked to whom and how much, DNS shows what they looked up, system/event logs carry OS-level attack indicators, application logs catch logins and SQL injection, and centralized auth logs tie identity together. Beyond the staples there's a long tail you pull in when the investigation needs it: VoIP/SIP, full packet captures, memory analyzers, scan output.

Every investigation I've run was bounded by one thing: what was actually being logged at the time. You cannot analyze a log you didn't collect, so the first job in security monitoring is deciding what to capture. The useful way to think about sources is by the question each one answers.

Network flow logs (netflow) answer "who talked to whom, and how much." They don't carry packet contents, but they show which systems communicated and the data volumes, which is exactly what you need to spot an exfiltration or a beacon. DNS logs answer "what did they try to reach": every external lookup, which is where a lot of C2 and phishing infrastructure shows itself before any payload moves. System and event logs are the OS-level record: security events and the indicators of an attack on the host itself. Application logs go a layer up: logins, data access, and the telltale signs

of things like SQL injection in the requests. And authentication logs, ideally centralized, tie identity across all of it, internal and external.



Then there's the specialized tail you reach for when a case demands it. VoIP and call-manager logs (SIP traffic) matter when the phone system is in scope. Full network traffic dumps and memory analyzers give you depth when flow logs and event logs aren't enough: the actual packets, the actual RAM. And vulnerability scan output is a log source in its own right, telling you what was weak at a point in time.

The practical lesson is that coverage is a decision you make *before* the incident, not during it. The gap you didn't fill six months ago is the blind spot you'll be staring into when it counts. Pick your sources deliberately, get them flowing to one place (which is the next problem), and remember that the cheapest log to have is the one you were already collecting.