SEAS-8414 CYBER ANALYTICS

01

Network Discovery and Asset Inventory

Descriptive cyber analytics for IoT/OT visibility, evidence, and bounded claims

Dr. Mallarapu

Analytics Taxonomy

12 chapters from descriptive discovery to autonomous remediation

CHAPTER TAKEAWAY

Chapter 1 owns the descriptive layer and constrains every later analytic stage.

ENRICHMENT VALUE

The deck uses the course architecture to show why undercounting here corrupts diagnosis, vulnerability work, and attack-path reasoning downstream.

Descriptive →	What is on the network? ARP, mDNS, SSDP, TCP probes	Ch 1: Discovery
Diagnostic →	What are these devices? nmap -sV, TLS, JARM, ONVIF	Ch 2: Enrichment
Detective →	What vulnerabilities exist? NVD/CVE, OpenVAS, Nuclei	Ch 3: Vulnerability
Predictive →	What attack paths are likely? BRS scoring, k-shortest paths	Ch 4: Attack Graphs
Prescriptive →	What offensive tests to run? RL agents, MDP, safety modes	Ch 5: Pentest
Autonomous →	Act on it — safely Digital twin, formal verify, closed-loop	Ch 6–12: Twin → Remediation

The Visibility Gap

You cannot protect what you cannot see

CHAPTER TAKEAWAY

IoT and OT failure often begins with assets that matter operationally but remain analytically under-observed.

ENRICHMENT VALUE

The lecture treats the gap as a strategic condition that later adversarial reasoning must explain, not just a missing count.

✔ CMDB Covers

- **Servers & VMs**
  Agent-managed, patched via SCCM/Ansible
- **Workstations**
  Domain-joined, EDR enrolled
- **Network Gear**
  Switches, routers, firewalls — SNMP managed
- **Cloud Instances**
  API-discovered, tagged, monitored

GAP

✘ CMDB Misses

- **IP Cameras**
  No agent, RTSP/ONVIF, default creds
- **HVAC Controllers**
  BACnet, 64MB RAM, no SSH
- **Badge Readers**
  Wiegand/OSDP, flat network access
- **Smart Displays**
  Android-based, no MDM, full TCP stack

The Common Attack Pattern

Common structural pattern across five documented incidents (see counterfactual analysis on next slide)

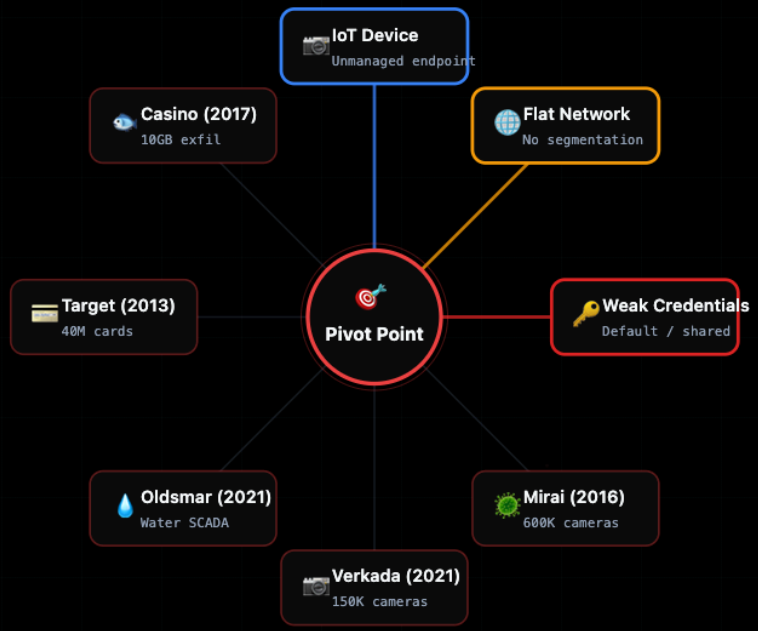
CHAPTER TAKEAWAY

Across incidents, overlooked edge devices become leverage because governance did not start with visibility.

ENRICHMENT VALUE

The visual compresses several famous incidents into one reusable structural skeleton students can apply beyond the named cases.

EVERY BREACH SHARES THREE FACTORS



Counterfactual Analysis: What Discovery Would and Would Not Have Prevented

Mapping five incidents to discovery coverage and orthogonal controls needed

CHAPTER TAKEAWAY

Discovery is necessary but not sufficient; it invalidates some preconditions and leaves others to orthogonal controls.

ENRICHMENT VALUE

The deck sharpens causal reasoning by making students separate visibility failures from credential, segmentation, and secret-management failures.

INCIDENT	ENTRY VECTOR	DISCOVERY ADDRESSES?	ORTHOGONAL CONTROLS NEEDED
Mirai (2016)	Default Telnet credentials on 600K cameras	Yes -- discovery + default cred check finds admin:admin	Credential rotation policy, network segmentation
Verkada (2021)	Super-admin credential exposed on public Jenkins	Partial -- discovers cameras, but not leaked credential	Secrets management, CI/CD hardening, MFA enforcement
Oldsmar (2021)	TeamViewer on SCADA with shared password	Yes -- discovers TeamViewer on OT subnet, flags remote access	OT/IT segmentation, jump-host architecture, MFA
Target (2013)	HVAC contractor pivot to POS VLAN	Yes -- identifies HVAC device on POS VLAN, topology anomaly	Micro-segmentation, vendor access controls, PAM
Casino (2017)	IoT thermometer cloud API to internal DB	Partial -- discovers unknown IoT device, but not API trust chain	Zero-trust east-west, API gateway, data-flow analysis

Why Traditional Scanners Fail at IoT

Five compounding factors that break Nessus/Qualys workflows

CHAPTER TAKEAWAY

Traditional enterprise scanners fail when their closed-world assumptions meet non-agentable, protocol-diverse, topology-bound fleets.

ENRICHMENT VALUE

The lecture translates that mismatch into an explicit design problem: what evidence sources substitute when the CMDB and agent model collapse.

01 📋 **Asset Inventory**
No agents on IoT -- 64MB RAM Hue bridge cannot run CrowdStrike

02 📡 **Protocol Diversity**
mDNS, SSDP, RTSP, ONVIF, MQTT, CoAP -- scanners only do TCP/UDP

03 🧬 **Identification**
Inference problem, not lookup -- MAC OUI tells vendor, not product

04 🔗 **Vulnerability Mapping**
CPE construction is fragile -- vendor:product:version must be exact

05 🌐 **Network Topology**
L2 vs L3 boundaries, VLANs, NAT -- need agents for remote subnets

Capability Comparison

Enterprise scanner vs. purpose-built IoT discovery platform

CHAPTER TAKEAWAY

No single method sees enough of the world to justify confidence alone.

ENRICHMENT VALUE

The slide turns methods into witnesses with different observational privileges instead of presenting them as interchangeable tools.

🏢 Nessus / Qualys

Discovery
ICMP ping + TCP port scan

Identification
Plugin-based, requires agent or creds

Protocols
TCP/UDP only, no mDNS/SSDP/ONVIF

IoT Coverage
Limited -- designed for servers/workstations

Topology
Flat -- scans from single vantage point

Cost
$3,500+/year per scanner

VS

🌊 IoT-Aware Platform

Discovery
7 techniques: ARP, mDNS, SSDP, TCP, fping, masscan, SNMP

Identification
Agentless fingerprinting: OUI + ports + banners + JARM

Protocols
mDNS, SSDP, RTSP, ONVIF, HTTP, TLS, JARM

IoT Coverage
Purpose-built for cameras, HVAC, badge readers

Topology
Distributed agents for remote L2 subnets

Cost
Open source, self-hosted

Measurement Boundary

Chapter 1 is about what enters the measured population, not about pretending the boundary disappears

CHAPTER TAKEAWAY

Discovery is about building D-hat, not pretending the true population is directly visible.

ENRICHMENT VALUE

The lecture names the boundary itself as an analytic object, which is what lets later claims about completeness stay honest.

OPEN-WORLD DISCOVERY

Measured Population

- Hosts with at least one justified observation
- Per-host provenance: ARP, router, SNMP, multicast, TCP
- Timestamps and method-specific caveats remain attached

Ambiguous Edge

- Silent-but-present devices
- Stale cache artifacts and phantom hosts
- Filtered responders and topology-confined peers

Outside Current Boundary

- Unseen routed segments
- Devices hidden by timing, privilege, or interface choice
- Anything not yet justified enough to enter inventory

SECTION 03

Architecture & Pipeline Overview

From subnet CIDR to risk report in seven phases

High-Level Architecture

Three clients, one API, external tooling via subprocess adapters

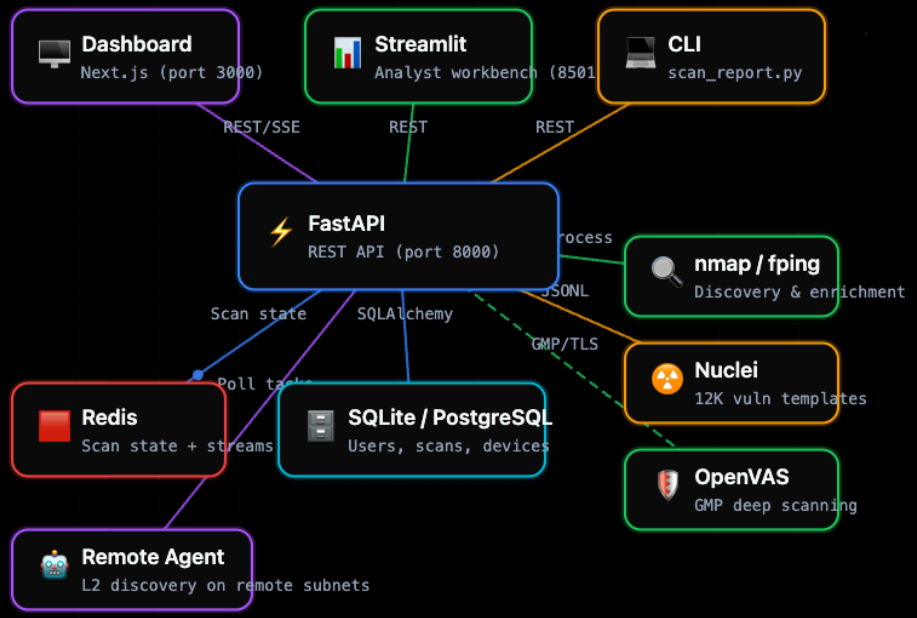
CHAPTER TAKEAWAY

The discovery system is a pipeline of observers, not a single scanner pass.

ENRICHMENT VALUE

The lecture exposes the control points where topology, concurrency, and evidence preservation become design choices.

BREAKWATER PLATFORM ARCHITECTURE



Scan Data Flow

From operator click to real-time dashboard updates

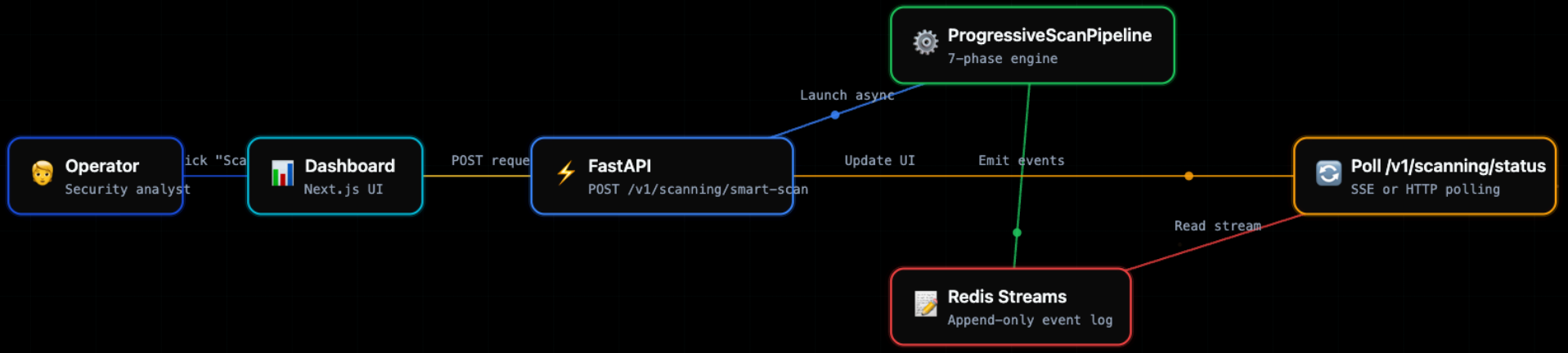
CHAPTER TAKEAWAY

Discovery output becomes useful when it moves from observation to persisted, queryable state quickly and legibly.

ENRICHMENT VALUE

Students see the operator-to-dashboard path as a living system rather than an abstract codebase reference.

END-TO-END DATA PATH



Seven-Phase Progressive Scan Pipeline

Each phase feeds the next with enriched data

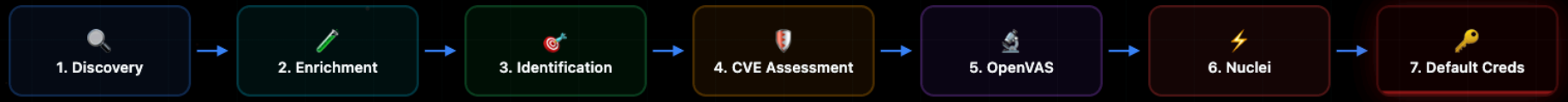
CHAPTER TAKEAWAY

Progressive discovery works because the system stages evidence and refinement instead of demanding perfect first-pass knowledge.

ENRICHMENT VALUE

The deck surfaces how sequencing itself controls what counts as early truth, provisional truth, and stronger later corroboration.

`PROGRESSIVESCANPIPELINE.RUN()`



Batch Concurrency Model

/16 subnet split into 256 /24 batches -- phases 1-4 run concurrently per batch

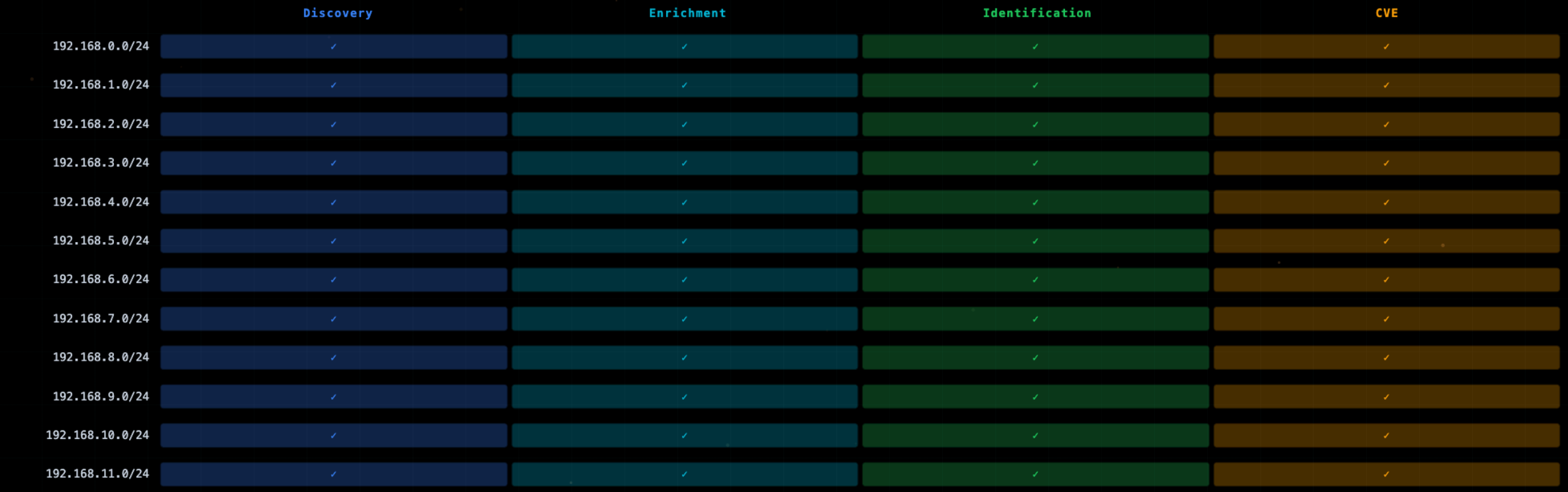
CHAPTER TAKEAWAY

Concurrency is a bounded quality decision, not a synonym for maturity.

ENRICHMENT VALUE

The lecture connects queue and worker design directly to traffic discipline, socket exhaustion, and trust in negative evidence.

`192.168.0.0/16` → 256 x /24 batches → `batch_concurrency=4`



Producer-Consumer Pattern

Bounded asyncio queues with None sentinel for end-of-stream signaling

CHAPTER TAKEAWAY

Streaming discovery improves operator usefulness because inventory forms before the entire run is complete.

ENRICHMENT VALUE

The slide turns architecture into an epistemic advantage: time-to-first-truth becomes part of the method.

progressive.py — PYTHON

```python
# Producer: discovery phase feeds enrichment queue
async def _discover_and_enqueue(self, batch: str):
    hosts = await self._discover_hosts_fast(batch)
    for host in hosts:  ← Backpressure: blocks if queue full
        await self.enrichment_queue.put(host)   # bounded queue  ← None = no more hosts coming
    await self.enrichment_queue.put(None)       # sentinel: end-of-stream

# Consumer: enrichment workers drain queue
async def _enrichment_worker(self, worker_id: int):
    while True:  ← Re-broadcast so all workers stop
        host = await self.enrichment_queue.get()
        if host is None:                        # sentinel received
            await self.enrichment_queue.put(None) # re-broadcast for peers
            break
        enriched = await self._enrich_host(host)
        await self.ident_queue.put(enriched)    # feed next phase
    ← maxsize=32 prevents memory blow-up
# Pipeline wiring
self.enrichment_queue = asyncio.Queue(maxsize=32)
self.ident_queue = asyncio.Queue(maxsize=32)
```

Running Case: Evidence Ladder

The point of the quiet device is not the address itself, but how evidence gets stronger without collapsing into premature identity

CHAPTER TAKEAWAY

The quiet device becomes more real through corroboration, not through premature naming.

ENRICHMENT VALUE

The lecture makes evidence strength visible stage by stage so students can watch a weak claim harden without crossing into Chapter 2.

CASE RECORD

192.168.86.42

d4:f5:47:xx:xx:xx

OPEN PORTS	VENDOR HINT
80, 443	Google Inc.
MDNS / SSDP	**CHAPTER 1 STATUS**
No useful response	Existence justified

The chapter never asks more of the record than the evidence can support.

1 **ARP cache**
Recent local IP-to-MAC state
Probably exists on local subnet

2 **Router / SNMP corroboration**
Gateway saw the same address too
Not just a single stale local artifact

3 **No multicast self-description**
No mDNS or SSDP value observed
Silence stays ambiguous, not exculpatory

4 **TCP connect on 80/443**
Reachable service endpoint
Existence is strong; identity still unresolved

Observability Depends on Vantage Point

Discovery results are properties of method plus position, not of method alone

CHAPTER TAKEAWAY

Observability is always a joint property of method and vantage point.

ENRICHMENT VALUE

The matrix shows how the same method changes meaning across laptop, gateway, VPN, and switch-adjacent contexts.

VANTAGE	LOCAL ARP	ROUTER ARP	MDNS/SSDP	ACTIVE PROBE	INTERPRETATION
Laptop on Wi-Fi	High	Medium	Medium	High	Great for local chatter, weak for routed blind spots
Gateway with credentials	Low	High	Low	High	Best witness for adjacent segments and whole-subnet corroboration
Collector through VPN	Low	Low	Low	Medium	Easy to misread as silence when the interface is simply wrong
Switch-adjacent OT box	High	Medium	Medium	Low	Favors passive methods when disturbance budget is tight

Topology Detection

Adapting discovery techniques to network position

CHAPTER TAKEAWAY

Discovery quality rises when the system first infers what kind of network it is standing in.

ENRICHMENT VALUE

The lecture treats topology detection as adaptive method governance rather than background plumbing.

AUTO

Detect subnet type automatically based on routing table

- All L2 on local
- Skip L2 on remote
- Smart fallback

```
BREAKWATER_SCAN_TOPOLOGY=auto
```

LOCAL

Force L2 techniques -- scanner is on the same broadcast domain

- ARP cache harvest
- mDNS browse
- SSDP M-SEARCH
- Router ARP table

```
BREAKWATER_SCAN_TOPOLOGY=local
```

REMOTE

Skip L2 -- scanning across routers where ARP/mDNS do not work

- fping/masscan only
- Extended TCP port list
- Pilot probe (7 IPs)

```
BREAKWATER_SCAN_TOPOLOGY=remote
```

SECTION 04

The Discovery Engine

Seven techniques across four network layers

Seven Discovery Techniques

Multi-layer approach ensures maximum host coverage

CHAPTER TAKEAWAY

Multiple techniques matter because each covers blind spots the others inherit.

ENRICHMENT VALUE

The lecture uses the seven-method picture to teach complementarity as a formal design principle.

1	L2	**ARP Cache Harvest**	Read OS ARP table -- zero traffic, passive discovery	`arp_adapter.py`
2	L2	**Router ARP Table**	Default gateway ARP cache via DHCP router adapter	`dhcp_router_adapter.py`
3	L2	**SNMP Multi-Router**	ipNetToMediaPhysAddress MIB across campus VLANs	`snmp_adapter.py`
4	Multicast	**mDNS Browse**	Zeroconf service discovery on 224.0.0.251:5353	`mdns_adapter.py`
5	Multicast	**SSDP M-SEARCH**	UPnP device discovery on 239.255.255.250:1900	`ssdp_adapter.py`
6	L3	**Fast Sweep**	masscan SYN sweep (root) or fping ICMP (non-root)	`masscan/fping subprocess`
7	L4	**TCP Connect Probe**	IoT ports (local) or 55-port extended list (remote) with pilot probe	`tcp_probe in progressive.py`

Observability Classes for Network Asset Discovery

Four classes ordered by intrusiveness, each with distinct coverage and failure modes

CHAPTER TAKEAWAY

Discovery itself has a defense-in-depth logic: different observational layers compensate for one another.

ENRICHMENT VALUE

The deck reframes layered observation as measurement resilience rather than just security jargon.

Passive Local

L2 / DATA LINK

Local broadcast domain only

METHODS

- ARP cache harvest
- DHCP lease tables
- MAC address tables (SNMP)

FAILURE MODES

- Cannot cross routers
- Misses silent devices
- Stale cache entries

Control-Plane

L2–L3 / MANAGEMENT

Spans routers via infrastructure queries

METHODS

- SNMP router ARP tables
- NetFlow/sFlow records
- DHCP server logs

FAILURE MODES

- Requires SNMP credentials
- Protocol support varies
- Management plane access

Low-Disturbance Active

L3–L4 / NETWORK–TRANSPORT

Any reachable subnet

METHODS

- ICMP echo (fping)
- TCP SYN sweep (masscan)
- mDNS/SSDP multicast

FAILURE MODES

- Firewall drops probes
- Rate limiting
- UDP-only devices invisible to TCP

Rich Application

L5–L7 / APPLICATION

Per-host, per-service deep probe

METHODS

- nmap –sV service detection
- HTTP banner / TLS cert
- ONVIF / RTSP / UPnP

FAILURE MODES

- Slow (seconds per host)
- Version string unreliable
- Protocol-specific parsers

Discovery Order: Passive to Active

Maximize coverage with minimal network impact

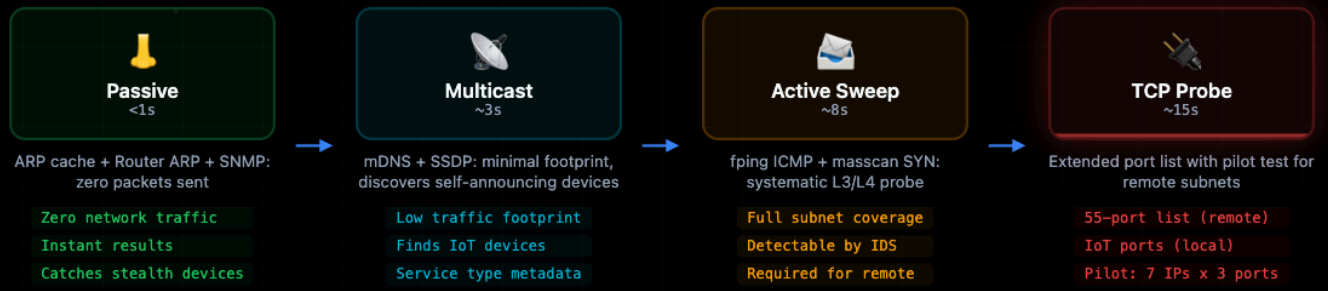
CHAPTER TAKEAWAY

Method ordering changes both packet cost and the interpretation of later evidence.

ENRICHMENT VALUE

The lecture shows that scan order is part of the analytic method, not just an implementation convenience.

INCREASING INTRUSIVENESS



ARP Cache Harvest Flow

Zero traffic generated -- reading the OS kernel's existing ARP table

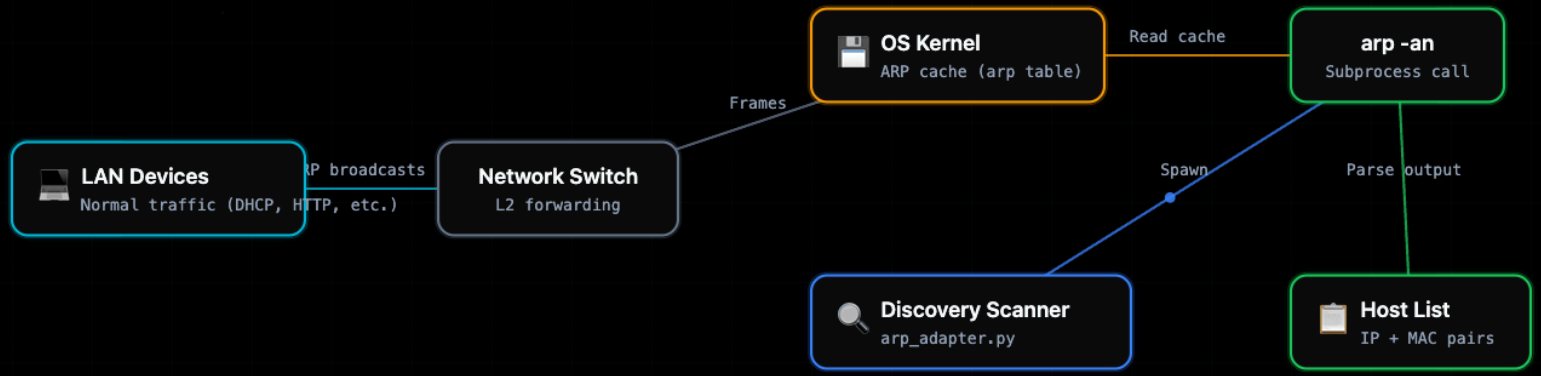
CHAPTER TAKEAWAY

ARP harvesting is powerful because it reads local network truth the system already holds.

ENRICHMENT VALUE

The animation makes the packet- and cache-level mechanics concrete enough that students can reason about what ARP can and cannot justify.

PASSIVE DISCOVERY VIA OS ARP CACHE



ARP Adapter Implementation

arp_adapter.py -- subprocess spawn, regex parse, subnet filter

CHAPTER TAKEAWAY

Small implementation choices like arp -an versus arp -a materially affect discovery usefulness.

ENRICHMENT VALUE

This slide teaches that engineering detail is not cosmetic; it shapes runtime cost and the cleanliness of evidence collection.

arp_adapter.py — PYTHON

```python
async def harvest_arp_cache(
    subnet: str | None = None,
) -> list[dict]:
    """Read OS ARP cache via 'arp -an' — zero network traffic."""
    try:
        proc = await asyncio.create_subprocess_exec(
            "arp", "-an",                    # -n: skip DNS reverse lookup   ← -n is critical: 4000x faster
            stdout=asyncio.subprocess.PIPE,
            stderr=asyncio.subprocess.PIPE,
        )
        stdout, _ = await asyncio.wait_for(
            proc.communicate(), timeout=5.0   # 5s timeout guard   ← Guard against arp hanging
        )
    except (FileNotFoundError, asyncio.TimeoutError):
        return []

    # Regex: ? (192.168.86.1) at aa:bb:cc:dd:ee:ff on en0
    pattern = re.compile(
        r"\((\d+\.\d+\.\d+\.\d+)\)\s+at\s+"
        r"([0-9a-fA-F:]+)"   ← Parse macOS/Linux arp output
    )

    hosts = []
    network = ipaddress.ip_network(subnet) if subnet else None
    for line in stdout.decode().splitlines():
```

ARP Cache Limitations

Understanding the boundaries of passive L2 discovery

CHAPTER TAKEAWAY

ARP is local, expiring, and incomplete, so its silence is weak negative evidence.

ENRICHMENT VALUE

The lecture turns ARP's limitations into a model for how to talk about any method honestly.

L2 only -- same broadcast domain

ARP operates at Layer 2. Routers do not forward ARP broadcasts across subnets.

→ *Cannot discover hosts on remote VLANs or across WAN links*

Cache entries expire (60s - 20min)

Default TTL varies by OS: Linux ~60s, macOS ~20min, Windows ~2min.

→ *Recently offline devices disappear from cache quickly*

Incomplete entries for inactive hosts

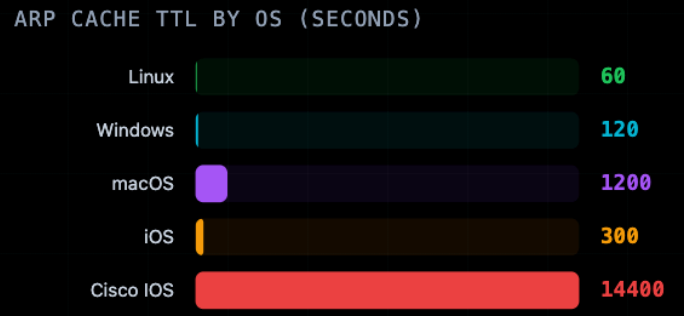
ARP marks entries as "(incomplete)" if no reply received within timeout.

→ *Breakwater filters these out -- only confirmed hosts reported*

No service or port information

ARP provides IP and MAC only -- no hostname, no service, no OS info.

→ *Enrichment phase (nmap -sV) is required for full device profiling*



Technique 2: Router ARP Table

The default gateway sees ALL devices -- not just those talking to the scanner

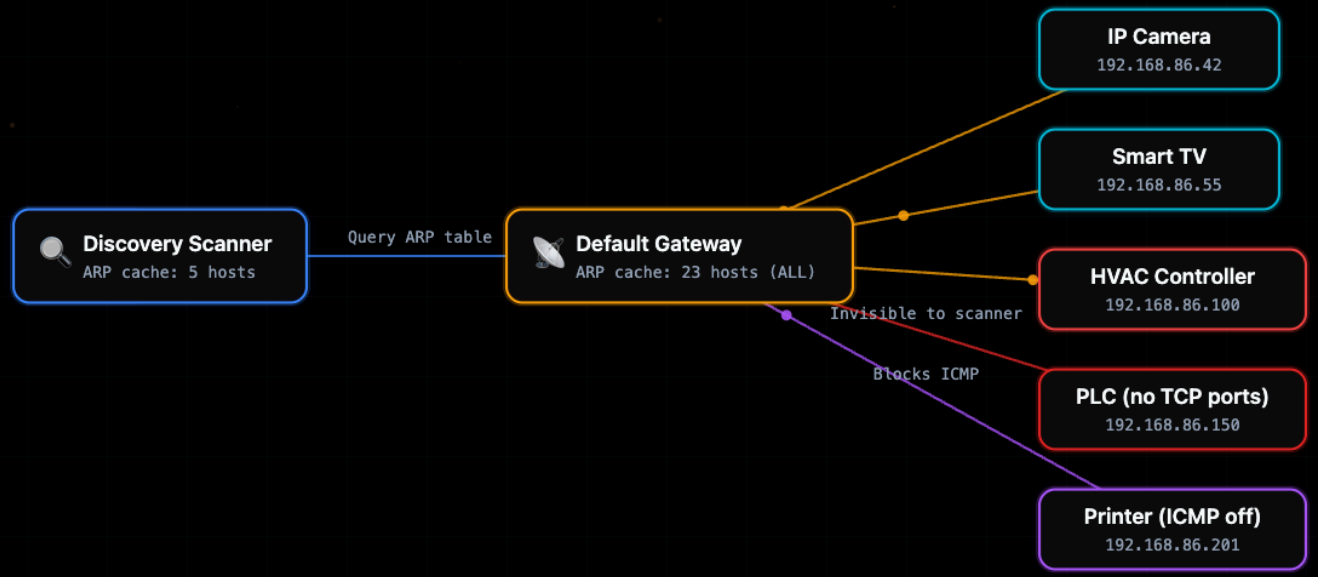
CHAPTER TAKEAWAY

When the host cannot see enough, the gateway often becomes the better witness.

ENRICHMENT VALUE

Students see how cross-subnet evidence can be gathered without immediately escalating to noisy broad sweeps.

GATEWAY ARP CACHE HAS COMPLETE VIEW



Technique 3: SNMP Multi-Router ARP Query

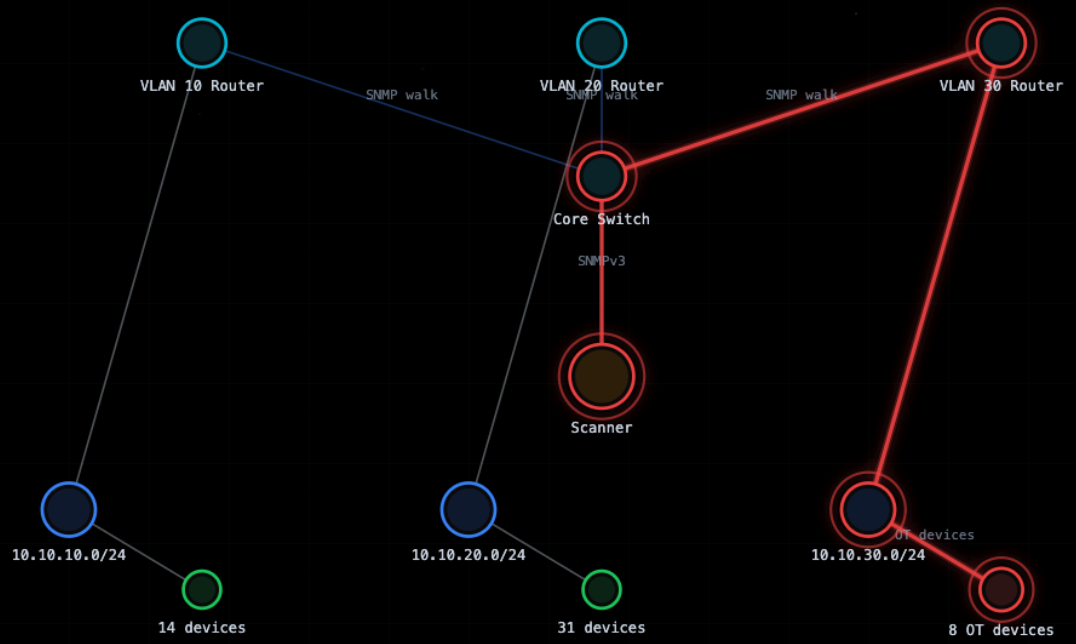
ipNetToMediaPhysAddress MIB -- campus-wide discovery across 50 VLANs via SNMPv3

CHAPTER TAKEAWAY

SNMP adds an infrastructure view that end-host methods cannot substitute for.

ENRICHMENT VALUE

The lecture links privilege and topology to observability by showing why asking the network can be more informative than asking the host.



Method Choice Follows Topology

The strongest discovery pipeline is not universal. It adapts to the kind of network it is standing in.

CHAPTER TAKEAWAY

The best discovery stack changes with local, routed, OT, and tunnel-confused environments.

ENRICHMENT VALUE

This slide makes method choice situational and comparative instead of treating one sequence as universally optimal.

Flat residential /24

ARP cache · mDNS · SSDP · fping · TCP connect

Local-link evidence is rich and cheap; active probing mainly fills silent gaps.

Routed enterprise VLANs

Router ARP · SNMP · pilot probe · TCP connect

The gateway becomes the better witness; link-local discovery alone undercounts.

Fragile OT / industrial segment

ARP cache · router telemetry · multicast only if normal · very limited active

Operational risk budget matters as much as nominal coverage.

VPN-confused remote collector

Interface validation · topology detection · router corroboration before multicast

Wrong interface choice can make a healthy network look empty.

Masscan Presweep Optimization

Sweep entire /16 first, then only scan /24 subnets with live hosts

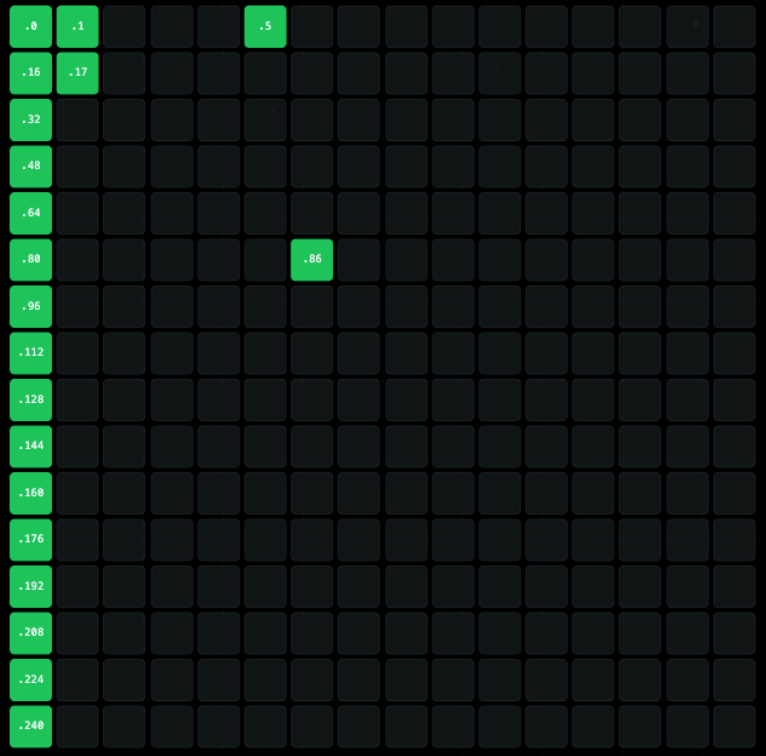
CHAPTER TAKEAWAY

Masscan is best used as search-space reduction, not as a source of final truth.

ENRICHMENT VALUE

The lecture turns a popular tool into a policy question: where should active budget be spent and where should it stop.

192.168.0–255.0/24 -- 256 SUBNETS



256	20	236
Total /24 subnets	Active subnets	Skipped (empty)

Without presweep: scan all 256 subnets

256 x 45s enrichment = ~3.2 hours for mostly empty subnets

With presweep: scan only 20 active subnets

7s presweep + 20 x 45s = ~15 min total (12x faster)

Passive-First Philosophy

Discovery ordered by intrusiveness -- gather what you can before sending probes

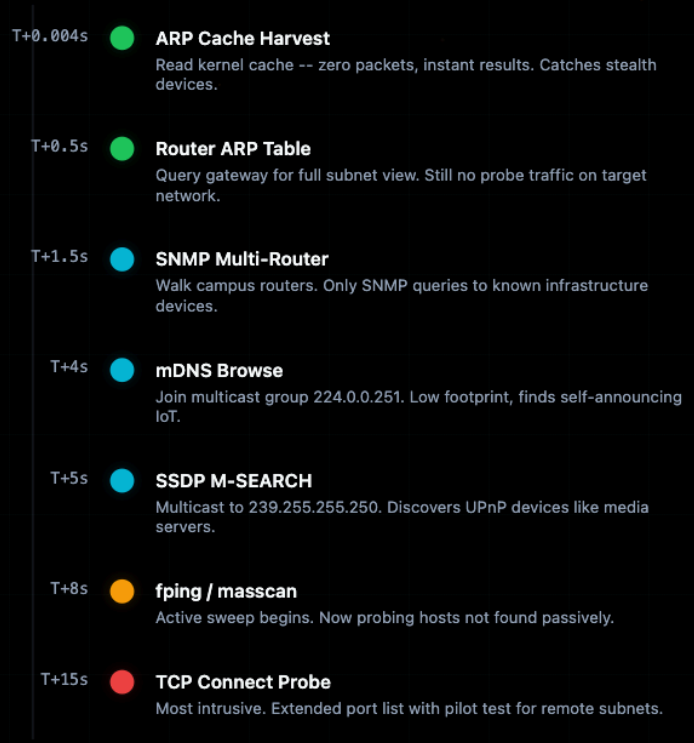
CHAPTER TAKEAWAY

Discovery should consume cheap evidence before emitting expensive traffic.

ENRICHMENT VALUE

The deck makes passive-first visible as both a safety rule and an interpretability rule.

DISCOVERY TECHNIQUE TIMELINE (PER /24 BATCH)



mDNS VPN Fix: _detect_lan_ip()

When VPN utun* interface hijacks multicast routing -- platform-specific engineering

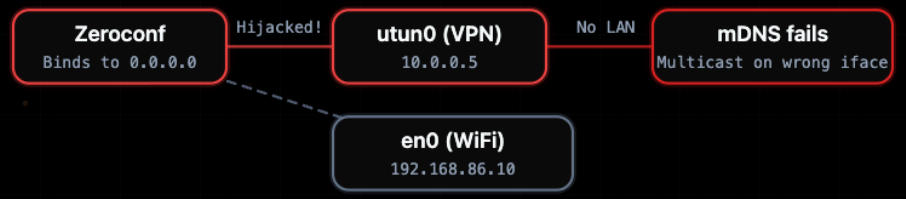
CHAPTER TAKEAWAY

mDNS is valuable but fragile because interface choice can silently invalidate the whole observation.

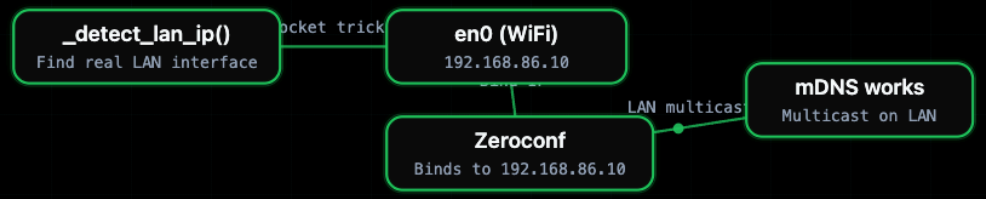
ENRICHMENT VALUE

The slide makes a quiet failure mode concrete enough that students stop treating “no response” as neutral evidence.

BEFORE: VPN HIJACKS MULTICAST



AFTER: EXPLICIT LAN BINDING



mdns_adapter.py — PYTHON

```python
def _detect_lan_ip() -> str:
    """Find LAN IP even when VPN utun* hijacks default route."""
    # UDP connect trick: doesn't send data, just resolves route
    s = socket.socket(socket.AF_INET, socket.SOCK_DGRAM)
    try:
        s.connect(("192.168.255.255", 1))  # LAN broadcast    ← Connect to LAN broadcast addr
        return s.getsockname()[0]          # Returns en0 IP   ← OS resolves to WiFi interface IP
    finally:
        s.close()

# Usage in mdns_adapter.py:
zc = Zeroconf(interfaces=[_detect_lan_ip()])    ← Force mDNS to correct interface
```

Passive-First Is a Risk Budget

The order is not aesthetic. It is how you keep discovery informative without turning the scanner into the incident.

CHAPTER TAKEAWAY

Probe choice is constrained by operational risk, especially in OT and clinical environments.

ENRICHMENT VALUE

The lecture gives students a risk-budget language for deciding how much coverage they can buy without becoming the incident.

Environment examples

Hospital / clinical
Prefer silence-preserving evidence first

Water / industrial
Probe budget is a safety decision

Residential / test lab
Can spend more active budget once passive clues thin out

Routed enterprise
Pilot probing trades time against broad blind search

Passive	ARP cache, router tables, SNMP reads	Near-zero disturbance
Low-disturbance	mDNS browse, SSDP M-SEARCH, limited control-plane queries	Small traffic cost, still environment-aware
Directed active	fping, TCP connect, pilot probes	Higher traffic cost, stronger negative-evidence ambiguity
Aggressive sweep	broad SYN scans, high-rate probing	Operational risk rises faster than insight on fragile segments

IoT Port Fingerprint

12 ports that identify 95% of consumer IoT devices

CHAPTER TAKEAWAY

TCP connect probing strengthens existence claims by tying them to reachable service endpoints.

ENRICHMENT VALUE

The deck grounds that logic in realistic IoT service surfaces instead of abstract “open port” language.

22 ADMIN SSH	23 ADMIN Telnet Cleartext protocol	80 WEB HTTP	443 WEB HTTPS
554 CAMERA RTSP Often unauthenticated	1883 IOT MQTT	5353 DISCOVERY mDNS	7000 APPLE AirPlay
8000 WEB Alt HTTP	8080 WEB HTTP Proxy	8443 WEB Alt HTTPS	8883 IOT MQTT/TLS

Pilot Probe

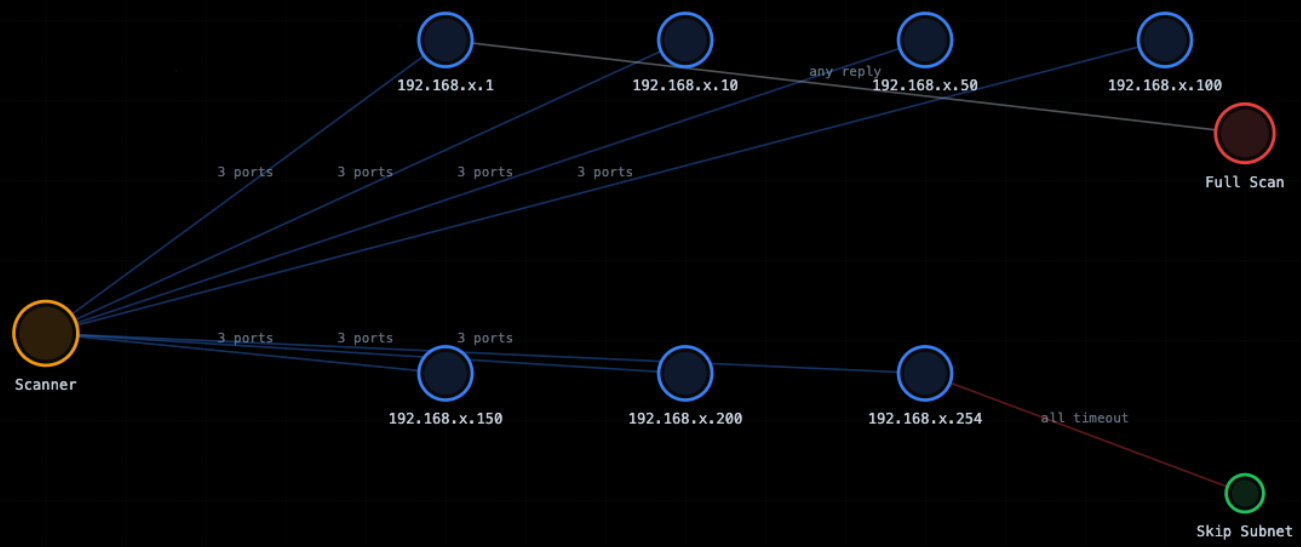
7 sample IPs x 3 ports -- skip entire subnet if unreachable

CHAPTER TAKEAWAY

Pilot probing is a budget-allocation strategy for large spaces, not a completeness guarantee.

ENRICHMENT VALUE

The slide prepares students to ask adaptive, policy-level questions about scan order and packet allocation.



Sample IPs: .1 .10 .50 .100 .150 .200 .254

Pilot Ports: 22, 80, 443

Saves on /16: ~12 min

Discovery Techniques Summary

Seven layers of progressive network discovery

CHAPTER TAKEAWAY

Each discovery technique differs in layer, scope, traffic cost, and strongest use case.

ENRICHMENT VALUE

The table turns the method stack into a comparative decision aid instead of a prose list.

#	TECHNIQUE	LAYER	SCOPE	TRAFFIC	BEST USE
1	**ARP Cache**	L2	Local	None (passive)	Instant neighbor map
2	**Router ARP**	L2/L3	Local	HTTP/SSH	Full subnet from router
3	**SNMP Router**	L2/L3	Multi-site	SNMP GET	Enterprise multi-VLAN
4	**fping/masscan**	L3	Any	ICMP/SYN	Fast alive detection
5	**mDNS Browse**	L3	Local	Multicast	Apple/IoT naming
6	**SSDP M-SEARCH**	L3	Local	Multicast	UPnP device detail
7	**TCP Connect**	L4	Any	TCP SYN+ACK	Port-based fingerprint

Discovery Results

IoT simulation: 22 devices on 172.30.0.0/24

CHAPTER TAKEAWAY

Method complementarity should show up empirically in host recovery and marginal contribution.

ENRICHMENT VALUE

The lecture makes students look for evidence that the ensemble is genuinely adding coverage rather than just repeating itself.

20/22
Devices Found

14s
Discovery Time

20
Unique MACs

2 devices missed
Powered-off devices invisible to all active techniques

DEVICES FOUND PER TECHNIQUE

Union of all techniques: **20 unique hosts** -- each technique contributes unique discoveries

The Host Record Is a Contract

Chapter 1 hands Chapter 2 an evidence-backed record, not an overconfident label

CHAPTER TAKEAWAY

Chapter 1 hands off an evidence-backed host record, not a flattened identity claim.

ENRICHMENT VALUE

The deck makes the record format itself part of the intellectual argument, not just an implementation artifact.

```
ip: 192.168.86.42
mac: d4:f5:47:xx:xx:xx
vendor_hint: Google Inc.
open_ports: [80, 443]
discovery_sources:
  - arp_cache
  - router_arp
  - tcp_connect
timestamps:
  first_seen: 2026-04-21T02:13:00-07:00
  last_seen: 2026-04-21T02:19:42-07:00
confidence_note: existence is well supported; identity remains
unresolved
```

`ip`
Stable network locator for follow-on work

`mac / vendor hint`
Link-layer anchor and weak vendor signal

`open_ports`
Observed services, not guessed identity

`discovery_sources`
Why the host claim exists at all

`timestamps`
When the evidence was actually observed

`confidence_note`
What remains unresolved and why

Nmap Caveats

When the most powerful tool can lose data

CHAPTER TAKEAWAY

Enrichment tools add value, but their outputs must remain bounded by Chapter 1's scope.

ENRICHMENT VALUE

The lecture shows how to borrow richer evidence without letting later-stage identity logic leak backward into discovery.

OS Fingerprint Overhead CRITICAL

-O --osscan-guess adds 15-30s per host. Combined with tight --host-timeout, nmap may drop the host entirely with ZERO results.

Host Timeout Trade-off HIGH

90s timeout is the compromise. Too low: drops slow-responding IoT devices. Too high: one stalled host blocks the enrichment queue.

Version Intensity MEDIUM

intensity 5 (default) sends ~40 probes. Intensity 9 sends 300+ but takes 3x longer. Most IoT devices respond at intensity 3.

Root vs Unprivileged INFO

Without root: TCP connect scan only (slower, more visible). With root: SYN scan + OS fingerprint + raw packet timing.

TLS Certificate Inspection

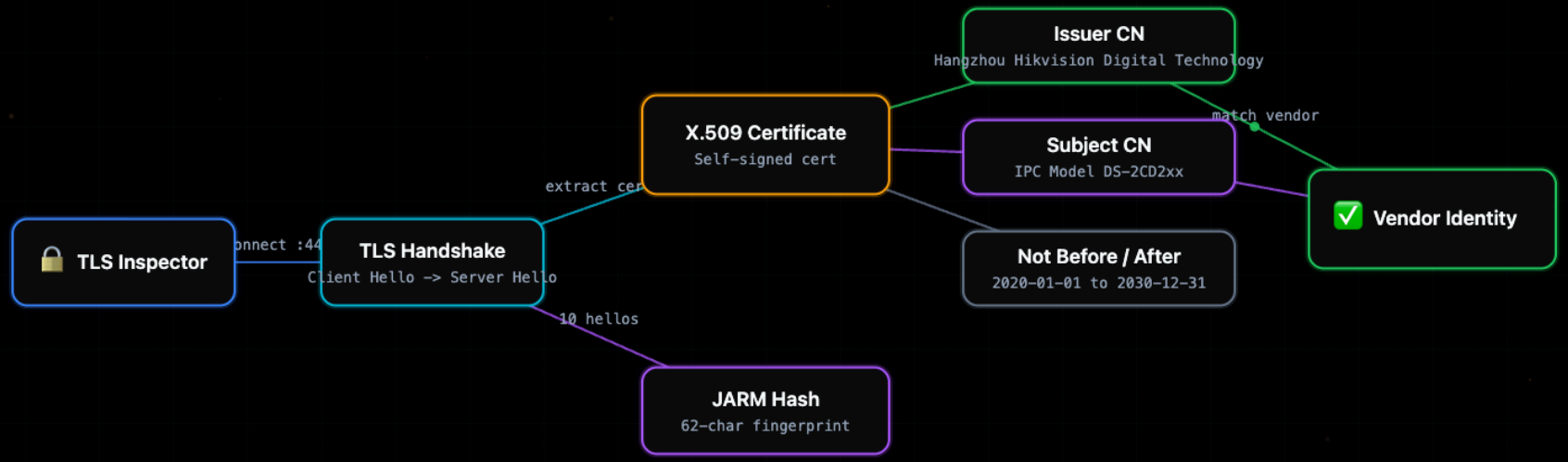
Self-signed certificates reveal manufacturer and model identity

CHAPTER TAKEAWAY

Some protocol surfaces already leak structured clues long before formal identification begins.

ENRICHMENT VALUE

The slide previews how later chapters will deepen identity while teaching students to keep current claims modest.



Issuer O
Hangzhou Hikvision Digital Technology Co., Ltd
Signal: Manufacturer

Subject CN
IPC Model DS-2CD2xx
Signal: Product model

Serial
Monotonic integer (not random)
Signal: Firmware vintage

Completeness, Soundness, and the Gap

A good Chapter 1 scorecard asks both 'what did we miss?' and 'what noise did we admit?'

CHAPTER TAKEAWAY

Completeness and soundness are different, and both are required for a trustworthy inventory.

ENRICHMENT VALUE

The lecture converts these into governance questions students can apply when reading any discovery claim.

Completeness

How much of the real population entered the inventory?

Low completeness means maneuver space remains for the attacker.

Soundness

How much of the inventory is actually real?

Low soundness pollutes every later stage with junk hosts.

Discovery Gap

What portion is still outside visibility right now?

A polished dashboard can still mask an under-observed network.

Capture-Recapture as an Audit Tool

When ground truth is unavailable, method overlap can tell you whether the discovery campaign is nearing saturation

CHAPTER TAKEAWAY

Hidden-population estimation can audit discovery saturation when direct ground truth is absent.

ENRICHMENT VALUE

The deck turns a compact formula into a methodological caution lesson rather than a magic completeness meter.

Worked example

Method A	ARP + router	18
Method B	mDNS + TCP	12
Overlap	Hosts seen by both	11
Estimate	N-hat = (18 x 12) / 11	19.6

If N-hat stays close to discovered count, the campaign may be near saturation.

If N-hat is materially larger, there may be a hidden population still outside the observed union.

This is not proof of completeness. It is a disciplined warning signal under explicit assumptions.

Best use: audit the shape of the search. Worst use: treating a small formula as a license to stop looking.

Adversarial Blind Spots

The attacker needs one overlooked foothold. The defender needs the blind spots to be small, named, and difficult to exploit quietly.

CHAPTER TAKEAWAY

The discovery gap is attacker maneuver space because silence and ambiguity can be exploited.

ENRICHMENT VALUE

The lecture explicitly brings adversarial reasoning into Chapter 1 without collapsing into offensive content.

Quiet protocol posture

Silence mDNS, SSDP, or ICMP while remaining reachable on a narrow management port.

Interface confusion

Let the defender point discovery at the wrong interface or tunnel and infer emptiness.

Timing and intermittency

Hide behind sleepy devices, short observation windows, or selective responses.

False-positive pollution

Exploit stale artifacts and noisy neighborhoods so the defender trusts the inventory less.

The Adapter Pattern

Consistent contract across all enrichment probes

CHAPTER TAKEAWAY

A modular adapter design makes method diversity manageable without flattening method-specific semantics.

ENRICHMENT VALUE

This slide links software design to epistemic clarity by showing why each method should keep its own evidence meaning.

adapter_template.py — PYTHON

```python
# Every enrichment adapter follows this pattern:
async def {protocol}_enrich(ip: str, **kwargs) -> dict:    ← Consistent async interface
    """Public async API — called from enrichment queue."""
    try:
        # Sync I/O wrapped in thread pool executor
        raw = await asyncio.to_thread(    ← Sync -> async bridge
            _sync_{protocol}_probe, ip, **kwargs
        )
        # Parse raw output into structured dict    ← Structured result contract
        return _parse_{protocol}_result(raw)
    except asyncio.TimeoutError:    ← Empty dict, not exception
        logger.warning(f"{protocol} timeout for {ip}")
        return {}  # Empty dict = graceful skip
    except Exception as e:    ← Log + continue pattern
        logger.error(f"{protocol} failed for {ip}: {e}")
        return {}  # Never propagate — other probes continue
```

ADAPTER IMPLEMENTATIONS

nmap_adapter.py
`subprocess.run()`

tls_adapter.py
`ssl.get_server_certificate()`

http_adapter.py
`httpx.AsyncClient.get()`

rtsp_adapter.py
`socket.connect() + send()`

onvif_adapter.py
`SOAP HTTP POST`

ssdp_adapter.py
`UDP multicast`

Same interface = easy to add new probes. Just implement the 3 functions.

Evidence Fusion, Calibration, and Controlled Propagation

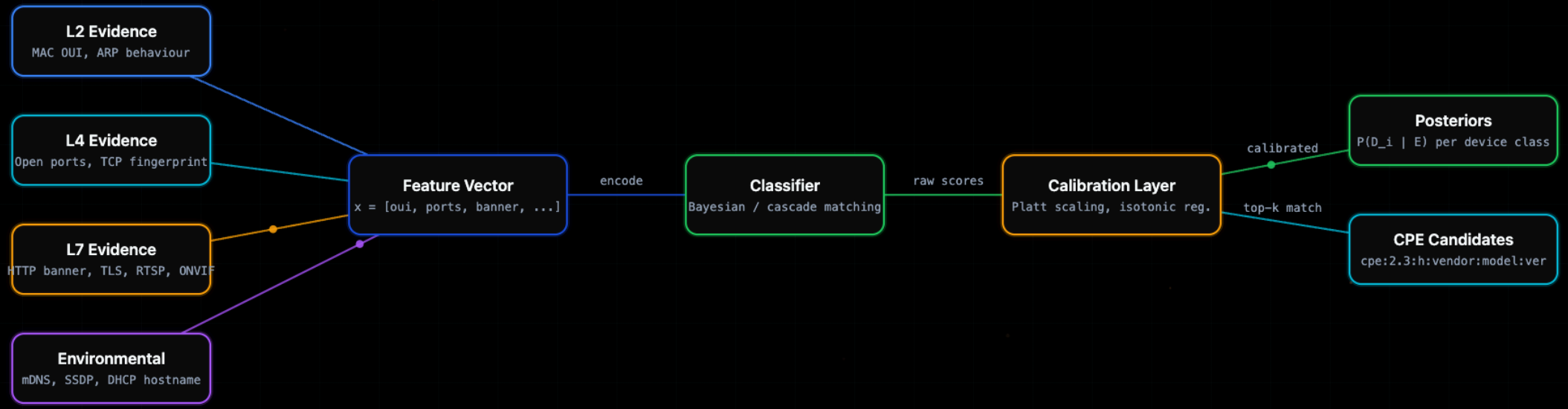
Four evidence categories flow through feature encoding, classification, and calibration to produce posteriors and CPE candidates

CHAPTER TAKEAWAY

Chapter 1 becomes trustworthy when multiple weak signals are combined without overstating what they prove.

ENRICHMENT VALUE

The lecture makes synthesis a disciplined merge problem rather than a vague intuition about "more data."



Confidence Scoring

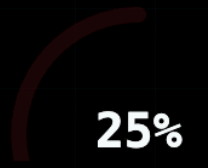
Each evidence layer increases identification certainty

CHAPTER TAKEAWAY

Confidence should rise with corroboration and remain visibly bounded when evidence is partial.

ENRICHMENT VALUE

The deck shows students how to talk about stronger and weaker host claims without drifting into theatrical certainty.



25%

OUI Only

20–30%

MAC vendor prefix match alone

55%

+ Hostname

50–60%

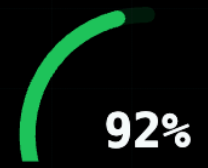
Hostname pattern adds context

70%

+ Open Ports

65–75%

Port signature narrows device type



92%

Full Match

85–100%

Service banners + product strings confirm

A disciplined discovery system never guesses a single identity. It accumulates evidence and reports the confidence level alongside the identification.

Chapter 1 to Chapter 2 Handoff

Existence is the hard question at the start. Identity becomes the hard question at the end.

CHAPTER TAKEAWAY

Chapter 1 stops at justified existence so Chapter 2 can ask the harder identity question honestly.

ENRICHMENT VALUE

The lecture uses the handoff itself to teach stage discipline and research hygiene.

Chapter 1 settles

- What enters the measured population
- Why each host claim exists
- Where the measurement boundary still leaks
- What later chapters may trust as evidence

Chapter 2 must answer

- What kind of device this actually is
- How banners, ports, certificates, and protocol traces cohere
- When identity confidence is high enough to act on
- Which product-level inferences are still weak or conflicting

Scan Timing Breakdown

Where time is spent in a typical full pipeline scan (~835s)

CHAPTER TAKEAWAY

Runtime is part of discovery quality when it affects operator usefulness and packet budgeting.

ENRICHMENT VALUE

The slide turns performance into a methodological variable rather than an engineering vanity metric.

Discovery ARP + fping + mDNS + SSDP + TCP

Enrichment nmap -sV, HTTP banners, TLS, JARM

Identification OUI + port sig + service match

CVE Lookup NVD API rate limits dominate

Default Creds HTTP, SSH, Telnet, RTSP, ONVIF

Nuclei ~12k templates, network tags

⚠ **Dominant Bottleneck: NVD API Rate Limits**

Without API key: 5 req/30s. With API key: 50 req/30s. CVE lookup takes 48% of total scan time.

Evaluation Methodology for Doctoral-Grade Research

Four pillars: ground-truth provenance, stratified datasets, ablation design, and calibration metrics

CHAPTER TAKEAWAY

False positives matter because they poison later reasoning, not because they merely clutter dashboards.

ENRICHMENT VALUE

The lecture ties soundness to trust, cleanup policy, and downstream analytic integrity.

Ground-Truth Sources

- CMDB-verified enterprise inventory
- Manual inspection (MAC, label, firmware)
- ONVIF/UPnP self-reported identity
- Vendor-confirmed test lab devices
- Crowd-sourced Fingerbank contributions

Dataset Strata

- Enterprise campus (managed switches)
- SMB flat network (consumer router)
- Industrial OT (Modbus/BACnet mix)
- Smart home (mDNS-heavy, NAT)
- Simulated lab (22 emulated devices)

Ablation Experiments

- Remove OUI prior -- measure accuracy drop
- Remove L7 banners -- isolate port-only ID
- Remove TLS/JARM -- test without crypto signals
- Single-technique vs. fusion comparison
- Vary confidence threshold (0.5 – 0.95)

Metrics

- Precision / Recall / F1 per device class
- Macro-F1 across all classes (>0.85 target)
- Top-k accuracy (k=1, 3, 5)
- ECE (Expected Calibration Error) < 0.05
- Brier score for probability quality

IoT Simulation Lab Architecture

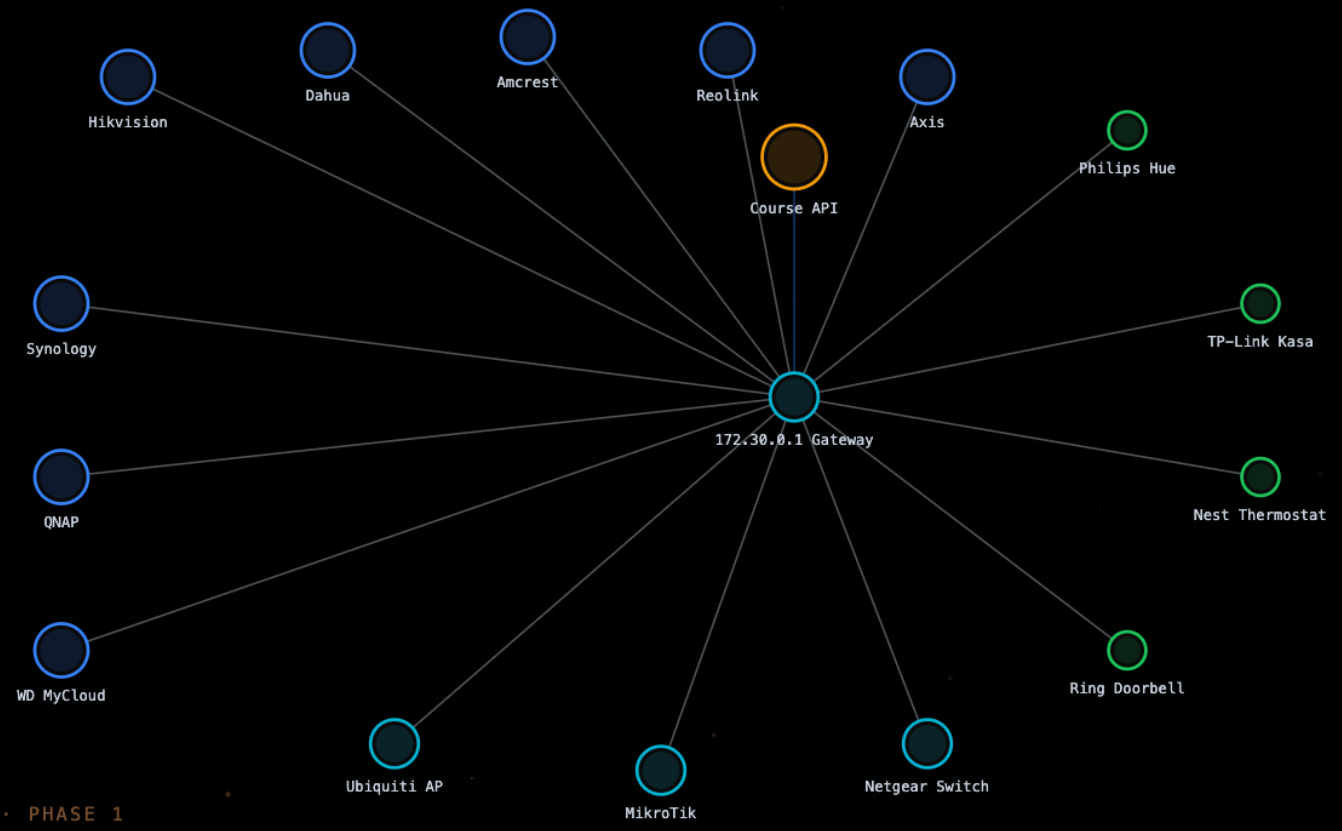
22 Docker containers on 172.30.0.0/24 (iotsim-net) with make iot-sim-up

CHAPTER TAKEAWAY

A good lab should expose discovery mechanics clearly enough that students can test claims rather than just follow steps.

ENRICHMENT VALUE

The slide bridges manuscript argument to practical experimentation, which is the right launch point for the lab phase.



Agent Architecture

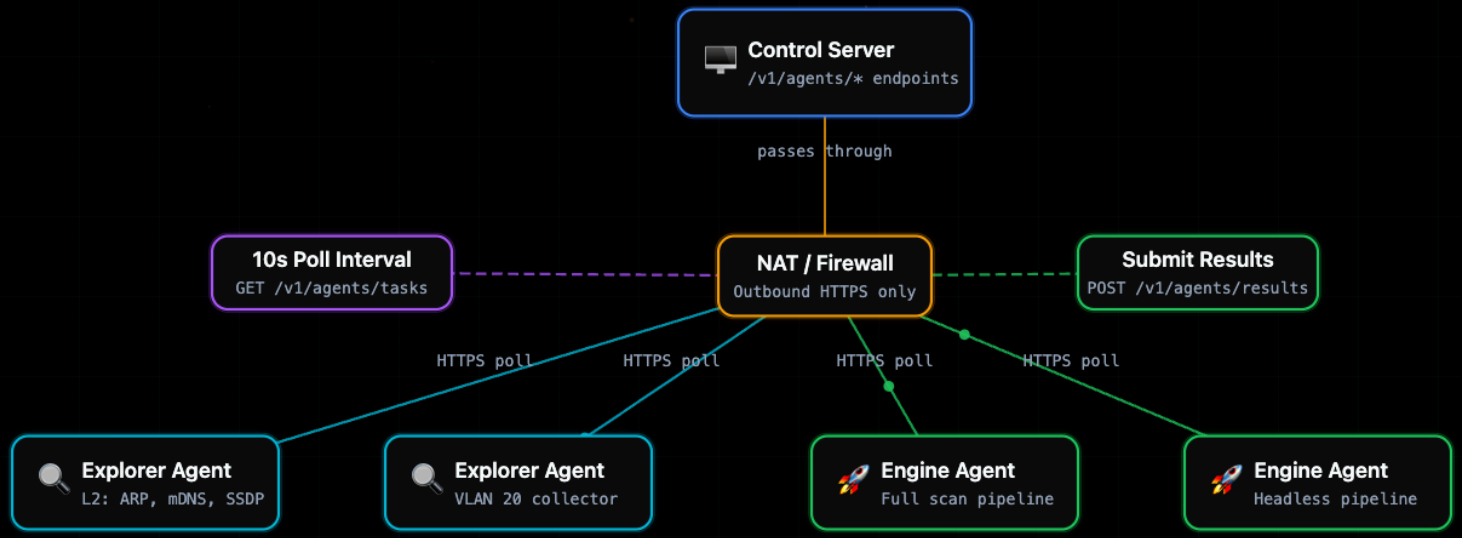
REST polling (not WebSocket/gRPC) — works through NATs and firewalls

CHAPTER TAKEAWAY

Distributed collection becomes necessary when the environment itself defeats single-vantage observation.

ENRICHMENT VALUE

The lecture broadens discovery from one scanner on one subnet to a realistic multi-vantage architecture problem.



Key Takeaways

Four essential lessons from Phase 1

CHAPTER TAKEAWAY

Discovery is a governance, measurement, and design discipline whose output is a defensible host inventory.

ENRICHMENT VALUE

The close synthesizes the lecture into a research agenda and prepares students to test the ideas in lab rather than merely remember them.

No Single Technique Finds Everything

ARP misses devices outside your VLAN. mDNS misses non-Apple. SSDP misses headless devices. TCP misses firewalled hosts. You need all seven discovery methods working together.

IoT Is Fundamentally Different

No endpoint agents, no patch management, factory credentials, 10-year lifecycles. Security requires compensating controls: segmentation, ACLs, monitoring.

Identification Enables CVE Assessment

Without fingerprinting a device type, you cannot construct a CPE string. Without a CPE, NVD lookup returns nothing. The pipeline order is not arbitrary.

Security Is Continuous

Scheduled scans, behavioral baselines, agent-based distributed coverage. A single scan is a snapshot — real security requires continuous assessment.