3AM Incident Survival Runbook

It’s 3AM.

You get paged.
Something is broken.
You don’t yet know what.

While it may feel random in the moment. It’s usually not.

The problem is rarely lack of tools.
It’s slow clarity in the first 15 minutes.

This runbook is for that window.

The first 15 minutes

1. **Acknowledge**
    - Ack the alert
    - Stop alert storms if possible
    - Take ownership
2. **Answer 3 questions fast**
    - What is broken?
    - Who is impacted?
    - Since when?

If you can’t answer this yet → don’t deep dive.

3. **Check recent change**
    - Deploys (last 30–60 mins or latest)
    - Config changes
    - Feature flags
    - Infra events
    - Third-party changes (CMS, payments, etc.)
4. **Stabilise before diagnosing**
    - Rollback
    - Scale up
    - Disable feature
    - Route traffic away
    - Revert the change

At 3AM, speed matters more than perfection.

5. **Narrow scope**
    - One service or many?
    - One region or global?
    - One endpoint or all traffic?

Triage patterns

Error spike after deploy

- Likely: bad release
- Action: rollback
- Verify: error drops

Latency increase (no errors)

- Likely: slow dependency or database
- Check: p95/p99, downstream calls
- Action: isolate slow path

CPU / memory saturation

- Likely: traffic spike or inefficient code
- Check: recent deploy, load pattern
- Action: scale or rollback

More triage patterns

Dependency timeout

- Likely: downstream service degraded
- Action: fail fast, reduce load, isolate

Database connection exhaustion

- Likely: connection leak or spike
- Check: active vs limit
- Action: restart / increase pool (temporary)

Queue backlog

- Likely: consumers stuck or slow
- Check: worker health, processing rate
- Action: scale consumers / fix blocker

CDN / cache issue

- Likely: cache miss or config change
- Check: hit ratio, origin load
- Action: restore caching / reduce origin pressure

Incident communication

Current impact: <who/what is affected>
Suspected area: <service / dependency / unknown>
Action in progress: <what you’re doing>
Next update: <time>

Example:

Current impact:
Checkout failures for EU users

Suspected area:
Recent payment service deploy

Action in progress:
Rolling back

Next update:
10 minutes

What not to do at 3AM

Most incidents get worse here.

- Don’t debug everything at once
- Don’t chase multiple theories
- Don’t ignore recent changes
- Don’t restart blindly
- Don’t add people without roles
- Don’t optimize before stabilizing
- Don’t assume the alert is correct

Closing

At 3AM, you don’t need perfect answers.

You need:

- fast clarity
- simple decisions
- reduced impact

Root cause can wait. Impact can’t.

More field notes: 3amsre.com