CF-Proxied Automation APIs — Whitelist BIC (INFRA-CF-001)

Rule: CF-Proxied Automation APIs — Whitelist BIC

ID: INFRA-CF-001
Priority: MEDIUM
Created: 2026-04-20
Related incident: LightRAG 46h outage (MC #8487 followup)

Rule

Svaki CF-proxied hostname koji servisira headless HTTP klijente (LightRAG, email-agent, pi-orchestrator, automation daemone, containerized services) MORA imati Cloudflare Configuration Rule koja disable-uje Browser Integrity Check (BIC).

Why

BIC filtrira requestove na osnovu User-Agent-a i fingerprint-a. Python `urllib` / `requests` / `httpx` default UA triggeruje block (HTTP 403, error code 1010) čak i ako je IP u Access bypass listi. BIC layer se evaluira PRIJE Access policies.

Ovo se ne vidi odmah jer:

- curl/wget imaju whitelisted UA → rade
- Browser testovi rade
- Samo python/node/go HTTP klijenti fail-aju

How to apply

Za svaki novi CF-proxied automation endpoint:

1. Identifikuj hostname (npr. `ollama.basicconsulting.no`)
2. Kreiraj Configuration Rule kroz CF API:

```
# Get zone ID first
export CF_ZONE_ID="<your-zone-id>"
export CF_API_TOKEN="<your-cloudflare-api-token>"

# Create rule
curl -X PUT
"https://api.cloudflare.com/client/v4/zones/${CF_ZONE_ID}/rulesets/phases/http_config
_settings/entrypoint" \
  -H "Authorization: Bearer ${CF_API_TOKEN}" \
  -H "Content-Type: application/json" \
  --data '{
    "rules": [{
      "action": "set_config",
      "action_parameters": {"bic": false},
      "expression": "(http.host eq \"HOSTNAME\")",
      "description": "Disable BIC for HOSTNAME (automation clients)",
      "enabled": true
    }]
  }'
```

3. Verifikuj: test query sa Python default UA — mora vratiti 200

Hostnames koji trebaju ovu zaštitu (live list)

- [x] ollama.basicconsulting.no (2026-04-20)
- [ ] lightrag.basicconsulting.no (has CF Access service token, mostly OK — but verify)
- [ ] Svi novi automation endpointi idu kroz ovu listu

Evidence

- CF Ruleset ID: 4fc2c122d04d4791a5d17409b097c510
- CF Rule ID: c5990f19f655441180ae886f4512de40
- Investigation: ~/system/evidence/lightrag-ingestion-investigation-20260420-215700.md

Gotchas

- Ne isključivati BIC globally — ostale zaštite ostaju aktivne
- Ne disable-uj na user-facing hostname-ovima (bilko.io, alai.no itd.) — tamo BIC štiti od botova
- Samo automation/API hostname-ovi

Technical Details

Problem:
LightRAG ingestion daemon (Python container) could not reach `ollama.basicconsulting.no` — all requests returned HTTP 403 with Cloudflare error code 1010 (Browser Integrity Check block). This was not visible in initial testing because:

- `curl`/`wget` from host machine → 200 OK (whitelisted UA)
- Browser tests → 200 OK (browser fingerprint passes)
- Python `urllib` from container → 403 (default UA blocked)

Root cause:
Cloudflare Access IP bypass policies evaluate AFTER Browser Integrity Check. Even with correct Access service token headers, BIC rejected the request based on User-Agent string before the Access policy could allow it.

Solution:
Configuration Rule that disables BIC for the specific hostname. Expression: `(http.host eq "ollama.basicconsulting.no")`. This allows automation clients through while preserving other Cloudflare security layers (WAF, DDoS, Access).

Time to incident:
46 hours from LightRAG Azure migration (2026-04-18) to detection (2026-04-20 21:00). Initial hypothesis: Neo4j memory pressure, NSG IP rotation, Azure network issue. All false. Real cause: CF security layer blocking automation client UA.

Fix time:
11 minutes from root cause identification to deployment + verification.

Generated by: ALAI, 2026

Revision #2
Created 2026-04-20 21:34:32 UTC by John
Updated 2026-05-31 20:06:21 UTC by John