**1) En qué consiste la comprobación**

**Objetivo:** detectar Message-IDs que sean **duplicados** (mismo Message-ID en varios correos distintos) o **fabricados/sospechosos** (formato raro o señales que indican generación automática/manipulación). Estas situaciones son relevantes porque un Message-ID repetido puede indicar que alguien clonó/duplicó mensajes, reusó cabeceras, o que hay manipulación en la evidencia; formatos raros o inconsistentes pueden indicar clientes no estándar, scripts de generación masiva o intentos de ocultación.

**Heurísticas que aplicamos (lista práctica):**

1. **Duplicado exacto:** mismo Message-ID aparece en más de un EML dentro del caso. **HECHO**
2. **Múltiples Message-ID en un mismo mail:** cabeceras Message-ID repetidas o múltiples cabeceras diferentes (anómalo). **HECHO**
3. **Formato inválido o incompleto:** falta de < >, ausencia de dominio, caracteres no permitidos. **HECHO**
4. **Patrones demasiado simples o secuenciales:** ej. 1@localhost, msg0001@example o 123456 — baja entropía. **HECHO**
5. **Dominio inconsistente:** dominio del Message-ID distinto al dominio desde el From o Return-Path (posible spoofing).
6. **Reuso entre remitentes distintos:** mismo Message-ID pero diferentes From/Return-Path.

necesitamos comparar **entre archivos**. Por eso:

* En process() guardamos en memoria (core.msgid\_register(...)) lo que vemos por archivo.
* En shutDown() hacemos la **comparación global** y creamos hits en **cada archivo implicado** cuando un Message-ID aparece en varios.

The "Message-ID:" field contains a single unique message identifier. The "References:" and "In-Reply-To:" field each contain one or more unique message identifiers, optionally separated by CFWS.

The "Message-ID:" field provides a unique message identifier that refers to a particular version of a particular message. The uniqueness of the message identifier is guaranteed by the host that generates it (see below). This message identifier is intended to be machine readable and not necessarily meaningful to humans. A message identifier pertains to exactly one instantiation of a particular message; subsequent revisions to the message each receive new message identifiers.

The message identifier (msg-id) itself MUST be a globally unique identifier for a message. The generator of the message entifier MUST guarantee that the msg-id is unique. There are several algorithms that can be used to accomplish this. Since the msg-id has a similar syntax to angle-addr (identical except that comments and folding white space are not allowed), a good method is to put the domain name (or a domain literal IP address) of the host on which the message identifier was created on the right hand side of the "@", and put a combination of the current absolute date and time along with some other currently unique (perhaps sequential) identifier available on the system (for example, a process id number) on the left hand side. Using a date on the left hand side and a domain name or domain literal on the right hand side makes it possible to guarantee uniqueness since no two hosts use the same domain name or IP address at the same time. Though other algorithms will work, it is RECOMMENDED that the right hand side contain some domain identifier (either of the host itself or otherwise) such that the generator of the message identifier can guarantee the uniqueness of the left hand side within the scope of that domain.

[ <https://datatracker.ietf.org/doc/html/rfc2822#section-3>

]

Me mira la estructura local@domain

Si no tiene Message-ID que salte la alarma

Es muy difícil porque cada MTA (mail transfer agent) hace su propio message-id.