RFC 6377 : DKIM And Mailing Lists

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 22 septembre 2011

Date de publication du RFC : Septembre 2011

https://www.bortzmeyer.org/6377.html

La sécurité, c'est compliqué. Ce n'est pas tout de définir des normes techniques comme DKIM, pour sécuriser le courrier électronique. Il faut encore étudier leurs conséquences pour tous les cas. Le courrier électronique ayant débuté il y a très longtemps, et n'ayant pas vraiment été spécifié dans tous les détails (son architecture n'a été officiellement décrite qu'en 2009, avec le RFC 5598 [1]), il représente un défi particulier pour tous ceux qui essaient de le sécuriser a posteriori. C'est ainsi que ce nouveau RFC s'attaque au cas particulier des listes de diffusion et étudie comment elles réagissent avec le mécanisme d'authentification DKIM.

D'abord, à quoi sert DKIM ? Normalisé dans le RFC 6376, DKIM permet de signer un message électronique. Sans lui, il est trivial de fabriquer un faux message (prétendant venir de votre banque et vous demandant de saisir votre mot de passe sur un formulaire Web) et il n'y a aucun moyen d'acquérir des certitudes sur ce message, où (presque) tout peut être faux. Avec DKIM, un **domaine** de courrier peut signer le message et ainsi en prendre la responsabilité. Si le message est signé par `mabanque.example`, vous pouvez être sûr, pas forcément de sa véracité, mais en tout cas d'un lien avec votre banque. Contrairement à PGP, conçu pour authentifier une personne, DKIM permet d'authentifier un domaine qui a une responsabilité dans l'envoi de ce message.

Mais les listes de diffusion posent des problèmes spécifiques (section 1). DKIM avait été conçu avec l'idée que le message n'était pas modifié en cours de route. S'il l'est, c'est considéré comme une attaque qu'il faut détecter. Or, les gestionnaires de listes de diffusion ne respectent pas ce principe. Par exemple, il est fréquent (bien que ce soit une très mauvaise idée, cf. le RFC 4096) que le sujet soit modifié pour y inclure le nom de la liste entre crochets. De même, l'ajout d'instructions de désabonnement à la fin de chaque message est commun. De telles modifications sont peut-être honnêtes (dans l'esprit de celui qui les a décidés, ce ne sont pas des tentatives de fraude) mais, pour un logiciel de vérification, elles suffisent à invalider la signature.

Si les intermédiaires qui relaient un message le laissaient rigoureusement intact, ce serait non seulement une bonne idée pour les utilisateurs, mais cela permettrait à DKIM de continuer à fonctionner. Comme, parmi les intermédiaires, les MLM ("*Mailing List Managers*") sont souvent les plus intrusifs, que peut faire DKIM pour gérer ce problème ?

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc5598.txt`

— Est-ce une bonne idée de signer avec DKIM un message envoyé à une liste de diffusion ?
— Est-ce que le MLM devrait vérifier ces signatures en entrée ?
— Est-ce que le MLM devrait retirer les signatures existantes ?
— Et/ou mettre la sienne ?

Ce RFC ne normalise pas de réponse officielle à ces questions : il explique le problème et pointe du doigt les compromis à faire. La recommandation générale est « dans le cas typique, le MLM a intérêt à signer les messages qu'il relaie, et les vérificateurs à la destination ont intérêt à tenir compte de cette signature ». Le lecteur attentif aura noté que cette recommandation laisse plein de questions ouvertes...

Avant de détailler les problèmes, la fin de la section 1 rappelle quelques éléments importants sur le courrier électronique et les MLM (voir le RFC 5598 pour une présentation générale de l'architecture du courrier et sa section 5.3 sur les MLM). Les MLM (les gestionnaires de listes de diffusion) n'ayant jamais fait l'objet d'une normalisation officielle (le RFC 5598 a été écrit longtemps après), leur comportement varie énormément. Lorsqu'un MLM fait une opération, il est bien difficile de trouver un RFC qui dit noir sur blanc s'il est en droit de la faire ou non. Une notion aussi simple que celle d'**identité** (qui est l'auteur du message) n'est pas évidente et DKIM a donc sagement décidé de ne **pas** lier le domaine signant (celui qui est indiqué par `d=` dans une signature DKIM) avec une des « identités » présentes dans les en-têtes du message. Ainsi, le domaine signant n'est pas forcément celui présent dans le champ `From:`. Il peut refléter, par exemple, un domaine qui a relayé le courrier et, en le signant, a affirmé une certaine responsabilité pour ce message. **Donc, si un MLM signe avec son domaine un message venu d'un autre domaine, cela ne pose pas de problème : c'est dans l'esprit de DKIM.**

C'est l'interprétation qui va poser problème. Si j'écris un message sur la liste smtp-fr@cru.fr <`https://listes.cru.fr/sympa/info/smtp-fr`> avec mon adresse professionnelle (`bortzmeyer@nic.fr`), qui est responsable ? L'« auteur », mentionné dans le champ `From:`, ou bien l'ex-CRU <`http://www.cru.fr/`>, qui gère le MLM, ou encore la personne mentionnée (ce terme est d'ailleurs malheureux) comme « propriétaire » de la liste smtp-fr (aujourd'hui, Christophe Wolfhugel). Selon le cas, un vérificateur sera plus intéressé par l'un ou l'autre des responsables.

Je l'ai dit, la normalisation des fonctions des MLM n'a été faite qu'a posteriori, n'est pas complète, et n'a pas été largement adoptée. C'est ainsi que des normes comme le champ `List-ID:` (RFC 2929) ou les en-têtes donnant les mécanismes de désabonnement (RFC 2369) ne sont que partiellement déployés.

Bref, que conseille ce RFC ? Avant d'en arriver aux conseils pratiques, en section 4 et 5, il faut encore bien se mettre à jour sur la terminologie (section 2) et les principes de fonctionnement des MLM (section 3). En section 2, il faut d'abord lire les RFC sur DKIM, le RFC 5585, qui sert d'introduction et le RFC 6376 sur la norme elle-même. Ensuite, notre RFC introduit quelques termes nouveaux comme « gentil avec DKIM » ("*DKIM-friendly*"), qui désigne un MLM dont les actions n'invalident **pas** les signatures DKIM. À noter que cela ne dépend pas que du logiciel mais aussi de sa configuration (si on lui fait ajouter un pied de page en bas de chaque message, les signatures portant sur le corps du message sont invalidées), et des options du signeur (si le MLM modifie le sujet mais que le signeur n'a pas signé le sujet, le MLM sera considéré comme gentil avec DKIM).

Alors, comment fonctionnent les MLM ? Le problème, note la section 3, est qu'il y en a plusieurs sortes. Les différents **rôles** qui décrivent les acteurs du système de messagerie (auteur, signeur, récepteur, etc) sont évidents lorsque le message est un envoi direct d'Alice à Bob. Mais le MLM perturbe ce bel ordonnancement théorique. Déjà, il y a quatre types de MLM :

— Les **alias**, qui transmettent le message sans autre changement que l'ajout de quelques en-têtes de trace (comme `Received:`). La section 3.9.2 du RFC 5321 les mentionne. Ils ne changent que l'enveloppe SMTP, pas le message (en tout cas pas dans les parties signées). Les alias se mettent typiquement en œuvre sur Unix dans les fichiers `aliases` (directives `alias_database` et `alias_maps` de Postfix). Ces MLM sont gentils avec DKIM et n'invalident jamais les signatures (on notera que c'est le contraire avec SPF, RFC 7208, puisque ce dernier authentifie l'enveloppe et pas le message).

— Les **réexpéditeurs** (*"resenders"*), qui reçoivent un message, lui font parfois subir quelques modifications, et renvoient ce message à la liste des abonnés. Ce sont les MLM typiques comme Mailman ou Sympa.
— Les **auteurs**. Ce sont les MLM qui créent le message, plutôt que de renvoyer un message qu'on leur a transmis. Pensez à une *"newsletter"*, ou au spam. Ils n'ont guère de problème avec DKIM puisqu'ils contrôlent toute la chaîne, depuis la rédaction initiale.
— Les **synthétiseurs** (*"digesters"*) qui prennent plusieurs messages reçus et les assemblent en un seul message, de type MIME `multipart/digest` (certains abonnés préfèrent lire ainsi). Cette synthèse est un nouveau message, mais composé à partir de messages pré-existants.

Dans les deux derniers cas, le MLM crée un nouveau message. Dans le second, le plus difficile, son rôle est moins clair. Si on veut approfondir ces nuances philosophiques, on peut consulter les sections 3.6.4 du RFC 5322 et 3.4.1 du RFC 5598.

Quels sont les effets exacts des MLM sur les messages signés ? La section 3.3 les détaille :

— Le sujet (`Subject:`) est parfois modifié, pour ajouter le nom de la liste entre crochets. Cette pratique est très contestée (personnellement, je ne l'aime pas du tout, car elle modifie le message pour rien : si on veut trier le courrier, il est plus simple de le faire automatiquement via Sieve ou un système équivalent) mais, comme le note le RFC, contestée ou pas, elle ne va pas disparaître du jour au lendemain. Comme le sujet est un des éléments les plus importants de l'en-tête (c'est souvent ce que l'utilisateur regarde en premier) et qu'il est donc recommandé de le signer, ce comportement des MLM pose un problème.
— En revanche, les en-têtes spécifiques aux listes de diffusion, comme le `List-ID:` du RFC 2929 ne posent a priori pas de problème, car ils sont rarement mis par les MUA et DKIM n'a pas de problème avec l'ajout de nouveaux champs.
— Mais il peut aussi y avoir des changements dans le corps du message. Ils sont parfois mineurs (par exemple l'ajout de trois lignes d'instructions de désabonnement à la fin de chaque message). Mais cela suffit à rendre la signature DKIM invalide. DKIM permet de limiter la portée de la signature du corps (par exemple, de ne signer que les N premiers octets), ce qui permet l'ajout ultérieur de contenu (au prix d'une sérieuse baisse de sécurité). Si les changements sont plus significatifs (par exemple traduire le message HTML en texte seul), alors la signature sera invalide quoiqu'il arrive.
— Enfin, certains MLM suppriment les fichiers attachés (présents dans une partie MIME du message), les remplaçant par exemple par l'URL d'un service d'hébergement où le fichier a été déposé. Cela cassera également la signature.

Bref, à l'heure actuelle, les MLM sont souvent méchants avec DKIM. Dans le futur, cela évoluera peut-être mais on ne peut pas en être sûr, notamment parce que DKIM n'est pas assez important pour que les programmeurs de MLM changent leurs pratiques juste pour lui.

Une fois ces pratiques des MLM étudiées, le RFC a deux parties de conseils. L'une est consacrée aux MLM qui ne connaissent pas DKIM. Cette section 4 fournit donc des conseils aux signeurs et aux validateurs pour le cas où le message est passé par un tel MLM. L'autre, la section 5, est composée de conseils aux MLM qui connaissent DKIM et veulent bien faire. Voyons d'abord la section 4. Si un message signé va passer par un MLM non-DKIM, que faire ?

Idéalement, si l'auteur sait qu'un message va passer par un MLM non-DKIM, il devrait décider de signer ou pas, selon le comportement du MLM. La signature est risquée car le MLM peut faire une des modifications citées plus haut, rendant la signature invalide et inquiétant les récepteurs. (En théorie, une signature invalide doit être ignorée - RFC 6376, section 6.1 - sauf si le domaine émetteur publie en ADSP une politique plus stricte.) Le problème de cette approche est que l'émetteur ne connaît pas forcément les détails de tous les MLM et autres logiciels par lesquels son message risque de passer. Les administrateurs système qui activent DKIM sur leur infrastructure vont donc devoir décider sans connaître toute l'information. Les politiques strictes de signature et de vérification n'ont donc de sens qu'en cas de contrôle complet de la chaîne entre l'émetteur et le récepteur (ce qui exclut les MLM). Ne

mettez donc pas une politique ADSP "*discardable*" si le courrier de ce flux de messages risque de passer par un MLM.

Cette absence d'information sur les MLM et autres logiciels intermédiaires touche également les vérificateurs. Une solution possible est d'exclure de la vérification les courriers passés par les listes. Ce serait évidemment un gros travail, sans compter la maintenance d'un tel système (lorsqu'une nouvelle liste apparait). Les vérificateurs ont donc tout intérêt à respecter l'un des principes de base de DKIM : ne pas jeter un message simplement parce que sa signature est invalide.

Et pour les MLM modernes et qui gèrent DKIM ? Que doivent-ils faire ? La section 5 expose, dans l'ordre du voyage du message, les recommandations qui leur sont destinées. Ajouter des en-têtes (comme `List-ID:`) ne pose pas de problèmes. C'est un comportement gentil (sauf dans le cas où ces en-têtes étaient déjà présents et signés, ce qui est rare).

Par contre, ajouter un texte à la fin du message, on l'a vu, casse la signature. L'IETF n'a pas de politique officielle à ce sujet et il n'existe pas de RFC qui dise noir sur blanc « "*Thou SHALL NOT change the text, even by a single byte*" ». Même si un tel RFC existait, rien ne dit que les MLM suivraient ses consignes d'autant plus qu'il n'existe pas d'en-tête standard pour des informations telles que les conditions d'usage de la liste. DKIM a normalement une technique pour ce cas : l'option `l=` qui indique la portée de la signature (en nombre d'octets). L'ajout de texte après la valeur indiquée par `l=` n'invalide pas la signature. Mais cette méthode est déconseillée : un de ses inconvénients est qu'un attaquant peut, lui aussi, ajouter ce qu'il veut au message, sans craindre d'être détecté.

Notre RFC suggère donc, plutôt que d'ajouter à la fin de chaque message un texte comme « Le spam et les remarques désagréables sont prohibées sur cette liste. Les règles d'utilisation complètes sont présentées en `http://www.example.com/lists/the-list/terms-of-use` », d'envoyer de telles informations de temps en temps via un message automatique. Certes, ce sera ennuyeux (pensez à tous les messages de Mailman le premier du mois) mais cela pourra être filtré facilement.

Le problème d'ADSP (RFC 5617) est encore plus fort. ADSP permet de publier dans le DNS des directives de validation du genre « Je signe tout et correctement, vous pouvez jeter tout ce qui n'est pas bien signé ». Si un domaine publie une politique ADSP "*discardable*", il interdit quasiment tout usage d'un MLM. Les seules solutions sont encore dans le futur. Par exemple, le RFC imagine des MLM qui connaissent ADSP et qui, lors de la réception d'une demande d'abonnement, testent la politique et, si elle est stricte, refusent l'abonnement. Mais la politique ADSP peut changer à tout moment, donc ce n'est pas une solution parfaite. Va-t-il falloir retester régulièrement ?

L'auteur d'un message signé est de toute façon toujours dans une position difficile : il ne sait pas ce qui va arriver à son message et il ne connaît pas les réglages des logiciels intermédiaires. La section 5.5 recommande donc de séparer le courrier en deux : celui qui sera forcément de bout-en-bout devrait être signé avec une identité et une clé différente du courrier général, qui sera peut-être relayé par un MLM. Ainsi, si `mabanquecherie.example` est le domaine d'une banque, le courrier envoyé directement au client ne devrait pas avoir le même `d=` que celui utilisé pour le travail quotidien des employés.

Et sur le MLM lui-même, s'il connaît DKIM, quelles sont les bonnes pratiques (section 5) ? La plupart des MLM procèdent à une forme d'authentification, souvent plutôt faible (par exemple, une vérification du `From:` vis-à-vis de la liste des abonnés). Avec DKIM, on pourrait avoir apparaître de nouvelles politiques, plus sûres (par exemple une exigence que le message soit signé et que le domaine indiqué par `d=` soit celui du champ `From:`). Évidemment, aujourd'hui, très peu de messages seront signés, donc le gérant du MLM va devoir décider quoi faire des autres. En tout cas, une vérification positive est

bon signe et le RFC demande que le MLM ajoute dans ce cas un en-tête indiquant une authentification réussie (RFC 7001).

Et que faire des signatures DKIM existantes, déjà présentes lorsque le MLM reçoit le message original ? Si le MLM est sûr de ne rien changer au message, il peut laisser ces signatures, qui apportent une information intéressante. Mais s'il modifie le message, il est sans doute plus prudent de les supprimer, pour éviter tout risque qu'un message arrive au destinataire final avec une signature invalide (ce que certains destinataires considéreront, en violation de la norme DKIM, comme une incitation à jeter le message).

Ensuite, il reste une autre décision à prendre pour le MLM : doit-il apposer sa propre signature ? Le RFC le recommande fortement, afin de bien marquer que le gérant du MLM accepte sa responsabilité pour le message qu'il transmet. Le risque est évidemment que certains destinataires interprètent cette signature comme signifiant que le message a été authentifié dès le début. Toutefois, DKIM en lui-même ne peut rien faire dans ce cas : il y aura toujours des gens pour mal interpréter une signature et lui attribuer un poids qu'elle n'a pas.

Continuons le trajet du message : nous approchons de la fin avec les conseils aux vérificateurs. En gros, le RFC demande que ceux-ci traitent les messages des listes comme n'importe quel autre message (dans l'état actuel du déploiement des RFC 2919 et RFC 2369, il n'y a pas de moyen fiable de savoir si un message est passé par un MLM).

Les listes de diffusion de l'IETF signent tous les messages sortants, depuis juillet 2011 <http://www.ietf.org/mail-archive/web/ietf-announce/current/msg09173.html>, avec d=ietf.org (les signatures originelles sont apparemment retirées). Barry Leiba a bien résumé ce que cela apporte : « *"What it does is allow you to assure yourself that the message was, indeed, from an IETF mailing list (well, from an IETF email server), and that it wasn't that someone tried to spoof that. That, in turn, allows you to confidently increase your trust that the message is not spam in proportion to your confidence in the IETF's spam-filtering capabilities."* ». Voici une telle signature :

```
DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/simple; d=ietf.org; s=ietf1;
        t=1311614276; bh=qLpIZcZ8XeP5xTrgVPRjnXlZjXWiz9DqXpACtarsL0Q=;
        h=Date:From:To:Subject:Message-ID:MIME-Version:List-Id:
        List-Unsubscribe:List-Archive:List-Post:List-Help:List-Subscribe:
        Content-Type:Content-Transfer-Encoding:Sender;
        b=ppseQobrat1rQ+Brsy2LSpMAA79YgaFJ7PK2EG1N4w0zS2IZBqDQiXYHJxG/wv4wl
        GOd42GtThBVxB5BmBhkTn8M1Rqz+ZhW2pLPlcI1zHcmLmJHLMt1wC6R3wiCi4bipVd
        CszNeb58HSYGNDQmvnW9dAxi38pL/kjunJTpmVT4=
```

Le RFC mentionnait à plusieurs reprises les MLM qui comprennent DKIM. À ma connaissance, aujourd'hui, il n'y a que Sympa dans ce cas. DKIM y a été introduit dans la version 6.1 <http://www.sympa.org/manual/dkim>. Quelqu'un connaît d'autres implémentations ?

Un bon article de fond en français sur la coexistence entre DKIM et les listes de diffusion est l'article de Serge Aumont aux JRES, « DKIM et les listes de diffusion <https://2009.jres.org/planning_files/summary/html/28.htm> ». Sinon, concernant les modifications d'en-tête par les MLM, un article toujours utile, quoique centré sur les modifications du Reply-To:, est « *""Reply-To" Munging Considered Harmful*" <http://www.unicom.com/pw/reply-to-harmful.html> ».

Merci à Serge Aumont pour sa relecture.