

Groupe de travail sur le pourriel

*Base de données canadienne
sur les pourriels*

Document de conception

Sous-groupe de travail sur les technologies et la gestion du réseau

mai 2005

Base de données canadienne sur les pourriels

Document de conception

Introduction

La stratégie du Groupe de travail canadien sur le pourriel fait appel à une approche multiple, axée sur un ensemble d'outils, pour lutter contre le pourriel. Cette stratégie repose sur une meilleure application des lois existantes, la prise de mesures par les secteurs d'Internet et de la technologie de l'information pour résoudre les problèmes de technologie des réseaux concernant le mauvais usage du courriel, de meilleures pratiques exemplaires adoptées par les fournisseurs de service et les entreprises qui utilisent le courriel à des fins commerciales légitimes, l'éducation et la sensibilisation du public ainsi que la collaboration à l'échelle internationale.

Le Groupe de travail examine notamment l'utilisation des lois et des procédures judiciaires actuelles pour lutter contre le pourriel, les exigences liées à l'application de la loi, les moyens de lutter contre les pourriels (par exemple, un mécanisme de plaintes, une plateforme afin d'aider les organismes d'application de la loi à établir des causes juridiques, etc.) et les répercussions en découlant; il encourage ces organismes à intenter des poursuites contre les polluposteurs. Par ailleurs, le Groupe de travail étudie les outils techniques dont on dispose et les approches qui s'offrent au secteur et au gouvernement pour réduire le volume de pourriels au Canada.

Les délibérations du Groupe de travail ont mis en évidence plusieurs difficultés qui l'empêchent de s'attaquer efficacement au problème du pourriel. Tout d'abord, on ne dispose au Canada d'aucun mécanisme bien établi pour recevoir les plaintes des entreprises et consommateurs canadiens qui reçoivent des pourriels. Les consommateurs se plaignent généralement à leur fournisseur de service Internet, et il n'existe aucun moyen de qualifier ou de quantifier l'ampleur actuelle du problème au-delà du niveau des différents fournisseurs de service Internet.

Deuxièmement, il est difficile de faire enquête sur les pourriels, car on doit posséder des compétences techniques pointues et trouver des exemples de plaintes ou d'infractions.

C'est pourquoi le Groupe de travail explore la possibilité de créer une base de données à laquelle les internautes pourraient envoyer copie des pourriels reçus dans leur ordinateur. Un organisme national répertorierait les pourriels signalés et les conserverait pendant une certaine période. Ces messages pourraient servir à des fins d'application de la loi et de recherche.

La Base de données canadienne sur les pourriels (ou « congélateur » selon le terme anglais) serait similaire à celles de la Federal Trade Commission des États-Unis (FTC), du groupe de travail américain contre l'« hameçonnage » (pêche aux données personnelles) et du projet canadien PhoneBusters portant sur les plaintes contre le télémarketing. Elle pourrait partager les renseignements avec ces organisations et d'autres organisations internationales à vocation analogue.

Objectifs de la Base de données

La Base de données canadienne recevrait et stockerait des exemples de pourriels signalés sur une base volontaire par les Canadiens. Ceux-ci disposeraient ainsi d'un mécanisme digne de confiance et efficace pour rapporter les pourriels et appuyer de ce fait les efforts déployés afin de faire respecter les lois canadiennes s'appliquant à ces messages et aux autres types de mauvais usage du courriel.

En plus d'être utile aux organismes chargés de l'application des lois canadiennes, la Base de données pourrait être mise à la disposition des organismes privés participant à la lutte contre les pourriels. Les groupes liés à l'application de la loi, les fournisseurs de service Internet, les organismes de recherche et les autres organisations compétentes, Industrie Canada par exemple, pourraient être au nombre des utilisateurs autorisés. L'information stockée dans la Base de données pourrait être utilisée en preuve dans les poursuites judiciaires, servir à l'analyse des données statistiques et aider à élaborer la politique du gouvernement du Canada dans le domaine.

La Base de données devrait apporter son utilité aux citoyens canadiens, aux organismes d'application de la loi et autres organisations gouvernementales, tout en préservant les droits à la vie privée tels que définis par la *Loi sur la protection des renseignements personnels et les documents électroniques* et d'autres lois canadiennes.

Conception technique

La présente section décrit la conception technique de la Base de données canadienne sur les pourriels afin de donner une bonne idée du fonctionnement du système. Elle se divise en deux parties :

- 1) Utilisation par les internautes canadiens
- 2) Structure de la Base de données

1) Utilisation par les internautes canadiens

Le but du « congélateur » est de fournir aux Canadiens un guichet unique pour acheminer leurs plaintes contre les pourriels ainsi que les exemples de messages reçus à leurs comptes de courriel personnel et d'affaires.

Afin de permettre aux plaignants de transmettre des pourriels au « congélateur », on mettrait à leur disposition un compte de courriel générique, en quelque sorte un dépôt central canadien pour les pourriels. Aux États-Unis par exemple, la FTC utilise l'adresse **spam@uce.org**. On pourrait aussi mettre en place un site Web doté d'une tribune électronique pour renseigner les utilisateurs et aider à recueillir tous les renseignements importants afin de signaler les pourriels en bonne et due forme au « congélateur ». Il faudrait mener certaines activités de communication et de sensibilisation pour que les Canadiens connaissent l'existence du mécanisme de plaintes et de la Base de données.

La création de différents comptes de courriel aiderait à classer les plaintes par catégories. Ainsi, `pourriels@congelateur.ca`, `hameconnage@congelateur.ca` et `potdemiell.pourriels@congelateur.ca` pourraient tous être des comptes types du « congelateur ». Il pourrait être utile d'attribuer à chaque personne signalant un pourriel une adresse de courriel unique pour permettre au système de pister les pourriels signalés par les différents utilisateurs. De cette façon, le système pourrait peut-être donner des conseils judicieux aux internautes les plus vulnérables aux pourriels (par exemple, modifier leur adresse de courriel).

2) Structure de la Base de données

Le « congelateur » nécessiterait une infrastructure en plusieurs niveaux pour assurer l'intégrité des messages et permettre la recherche indexée, le stockage et le chiffrement de l'information ainsi que la préservation des éléments de preuve.

Intégrité des messages

Il faudrait préserver l'intégrité des courriels en créant une « étiquette » pour chaque message. En plus de fournir un moyen de prouver l'intégrité des données, ces étiquettes indiqueraient la date et l'heure ainsi que d'autres métadonnées (données décrivant le message). Le message proprement dit ne devrait être ni modifié ni filtré. Il devrait être accepté par la Base de données même si le serveur assurant la connexion n'utilise pas le « Simple Mail Transfer Protocol » (SMTP). Lorsque l'on récupère un message dans la Base de données, il faudrait s'assurer de son intégrité. Le système devrait enlever les pièces jointes contenant des virus.

Le système devrait connaître les listes noires et blanches et devrait marquer les métadonnées en leur attribuant une cote selon le protocole Internet (IP). Cette façon de procéder permettrait également de purger ultérieurement les messages envoyés pour perturber la Base de données.

Recherche indexée

La réussite de la mise en œuvre de la Base de données reposerait sur la recherche indexée. Un certain prétraitement des messages serait nécessaire pour réduire le nombre de requêtes et en améliorer l'efficacité. À l'étape de la planification, on devrait se pencher sur l'indexage et la recherche. L'administrateur de la Base de données devrait collaborer avec la FTC des États-Unis afin de connaître les recherches les plus fréquentes et la configuration de son index. Les clés primaires de l'index devraient être uniques, et il faudrait utiliser les clés figurant dans l'en-tête de chaque message pour déterminer s'il s'agit bel et bien de clés uniques. Au besoin, on pourrait utiliser une combinaison de champs afin de créer le champ unique nécessaire. Il faudrait classer ou coter les messages au moment de la réception pour déterminer leur degré de « pollution » et la catégorie de courriel dont ils font partie.

Stockage

Les messages devraient être stockés sur un réseau redondant de disques indépendants pour éviter que la défaillance éventuelle d'un disque unique ne provoque une interruption de service ou la perte ou la corruption de données. Il faudrait aussi utiliser un fichier de journalisation afin

d'améliorer l'intégrité des données.

Sécurité

La sécurité de la Base de données devrait permettre d'utiliser le protocole TLS pour le chiffrement du serveur de passerelle à passerelle. On ne devrait pas tenir compte du logiciel client peu importe qu'il utilise le programme de chiffrement PGP ou le protocole MIME sécurisé, car les pourriels ne renferment pas de données de nature délicate.

Accès

Les mégaétiquettes du système devraient permettre de contrôler les droits d'accès à des fins de sécurité et de pistage. Il faudrait aussi un mécanisme d'enregistrement pour suivre les recherches et limiter l'accès à la Base de données.

L'accès à la Base de données devrait être limité selon la priorité et l'incidence sur la performance du système. On devrait envisager d'avoir recours à un programme de gestion des ressources du système, afin d'éviter que des recherches de grande importance n'utilisent toutes les ressources.

Le Groupe de travail recommande d'exécuter plusieurs sous-programmes de post-traitement des données, entre autres :

- 1) produire des rapports sur le volume de pourriels, les adresses URL, les types de pourriels, la fraude, les sources et les destinations physiques ciblées
- 2) créer des notes et des rapports personnalisés en fonction des seuils définis par les organismes d'application de la loi
- 3) saisir l'information relative au site lorsque le compte constitue un « pot de miel » ou qu'il s'agit d'un compte-leurre.

Plusieurs adresses URL envoyées dans des pourriels renferment une chaîne spéciale de caractères permettant d'identifier le destinataire du message. Afin d'éviter que des consommateurs légitimes ne soient identifiés avec des polluposteurs, il est primordial que la Base de données n'exécute pas de recherche automatique des adresses URL au moyen d'un robot pour capturer ces sites Web à des fins de preuve.

Dans le cas de pourriels signalés en vrac par des fournisseurs de services Internet provenant de domaines légitimes, lorsque la victime visée est un pot de miel ou un compte-leurre, l'équipe de mise en œuvre de la Base de données pourrait envisager de procéder à rebours afin de trouver le site en question.

Estimation des coûts

Les coûts du système envisagé varieraient grandement en fonction des décisions stratégiques, notamment en ce qui concerne l'admissibilité des pourriels signalés, les possibilités de partenariat et le contrôle de l'accès. Les coûts indiqués dans le présent document sont fondés sur

les hypothèses de base suivantes.

- Un programme de partenariat public-privé fournirait la plus grande partie de la main-d'œuvre nécessaire pour développer le logiciel (le Groupe de travail reconnaît que cette hypothèse n'est peut-être pas réaliste).
- La planification, l'élaboration des politiques et la supervision du fonctionnement seraient assurées par un comité d'examen formé de bénévoles (non rémunérés), dont la composition serait déterminée selon un quota attribué aux différents types d'organisations (par exemple, gouvernementales, commerciales et non gouvernementales).
- Dans certains cas, les pourriels seraient utilisés (à l'aide d'un piège à pourriel par exemple).
- Les organisations compétentes auraient accès à la Base de données, sous réserve de l'approbation du comité d'examen. Cet accès ne serait pas limité aux organismes d'application de la loi.
- Le personnel directement affecté aux opérations serait partagé avec un centre de traitement existant.

Budget (approximatif)

- Achat du matériel : 426 000 \$ pour la première année et 158 000 \$ par an pour les années ultérieures.
- Location du matériel : 470 000 \$ pour la première année et 230 000 \$ par an pour les années ultérieures.

Il serait possible de réduire les coûts de 10 000 \$ dans les deux cas si aucune capacité supplémentaire n'était nécessaire en matière d'accès au réseau.

Une estimation détaillée des coûts est présentée à l'annexe I du présent document.

Droit de propriété et gestion

Afin de mettre en œuvre l'approche multiple qu'il a élaborée et utilisée jusqu'à présent et de fournir un mécanisme de coordination continue pour appliquer ses recommandations, le Groupe de travail sur le pourriel explore la possibilité de mettre sur pied un centre de coordination, ou un centre canadien de lutte contre les pourriels, qui pourrait piloter le fonctionnement de la Base de données canadienne sur les pourriels.

L'organisme central pourrait être intégré à une organisation gouvernementale existante (par exemple, Industrie Canada, le Bureau de la concurrence, la Gendarmerie royale du Canada, d'autres services de police ou le Conseil de la radiodiffusion et des télécommunications canadiennes), à un nouvel organisme ou à un partenariat public-privé. La Base de données canadienne sur les pourriels pourrait ainsi faire partie d'une entité appelée à gérer le système canadien de traitement des plaintes concernant les pourriels.

Annexe I – Estimation détaillée des coûts

Les coûts du système envisagé varieraient grandement en fonction des décisions stratégiques, notamment en ce qui concerne l'admissibilité des pourriels signalés, les possibilités de partenariat et le contrôle de l'accès. Les coûts indiqués dans le présent document sont fondés sur deux séries d'hypothèses.

Hypothèses de base :

- Un programme de partenariat public-privé fournirait la plus grande partie de la main-d'œuvre nécessaire pour développer le logiciel (le Groupe de travail reconnaît que cette hypothèse n'est peut-être pas réaliste).
- La planification, l'élaboration des politiques et la supervision du fonctionnement seraient assurées par un comité d'examen formé de bénévoles (non rémunérés), dont la composition serait déterminée selon un quota attribué aux différents types d'organisations (par exemple, gouvernementales, commerciales et non gouvernementales).
- Dans certains cas, les pourriels seraient utilisés (à l'aide d'un piège à pourriel par exemple).
- Les organisations compétentes auraient accès à la Base de données, sous réserve de l'approbation du comité d'examen. Cet accès ne serait pas limité aux organismes d'application de la loi.
- Le personnel directement affecté aux opérations serait partagé avec un centre de traitement existant.

Hypothèses détaillées :

- Le système pourrait avoir un débit de 500 000 messages par jour (y compris ceux signalés en lots modestes) représentant en moyenne 10 kbits chacun, soit 5 Go par jour en plus du trafic de service (par exemple, les fichiers de journalisation, l'indexage, les rapports), ce qui représenterait un total de 1 Go par jour ou 2 To par an.
- Le système utiliserait du matériel de secours de base (c'est-à-dire des copies de sauvegarde des données sur DVD inscriptible ou bande magnétique).
- Les données seraient conservées en ligne pendant deux ans.
- La supervision serait assurée par un comité d'examen non rémunéré, qui s'occuperait de la planification et de l'examen des demandes d'accès.
- Le système partagerait un centre de traitement et de soutien ainsi que les installations d'accès à Internet (c'est-à-dire qu'il ne s'agirait pas d'une entité autonome).
- L'accès serait assuré par des installations sécurisées, des navigateurs classiques et une infrastructure d'authentification normalisée, sans accès personnalisé ou infrastructure de réseau étendu privé.
- Un partenariat public-privé bénévole permettrait d'obtenir des fonds supplémentaires pour couvrir les coûts de développement des logiciels.
- Les coûts liés au matériel varieraient grandement selon que l'on inclut le soutien ou la maintenance dans le cadre d'un accord de location ou que l'on achète le matériel en plus d'assumer le temps et les fournitures nécessaires pour l'installation et le soutien. Les

deux estimations sont présentées à la fin de la présente annexe. Les coûts de location sont basés sur des contrats similaires conclus dans l'industrie.

- Dans la mesure du possible, on utiliserait les logiciels provenant de la FTC pour réduire les coûts de développement et intégrer des logiciels de gestion de Base de données et d'indexage dont les coûts sont connus et la viabilité éprouvée.
- Les services juridiques (concernant par exemple la protection des renseignements personnels, la présentation des pourriels, les politiques d'accès et les conseils juridiques proprement dits) seraient assurés par les organismes parrains (par exemple, le Commissariat à la protection de la vie privée et Industrie Canada).

Coûts de mise en œuvre et de maintenance

Matériel et supports

- Deux PC serveurs de milieu de gamme utilisant la plateforme Linux :
 - location (maintenance regroupée) : 24 000 \$ par an, ou
 - achat : capital de 30 000 \$; 6 000 \$ par an pour la dépréciation, le remplacement et le soutien.
- Réseau redondant de disques indépendants de 5 To (durée de conservation de deux ans) :
 - location : 60 000 \$ par an, ou
 - achat : capital de 10 000 \$; 5 000 \$ par an pour la dépréciation, le remplacement et le soutien.
- Supports d'archivage : 5 000 \$ par an.

Infrastructure et réseautage

- Obtention de la plus grande partie de l'infrastructure (par exemple, le système de noms de domaine [DNS], le protocole de synchronisation réseau [NTP], le coupe-feu, les dispositifs de sécurité) grâce au partage des installations : 1 000 \$ par an pour les frais accessoires.
- Connexion réseau (5 Go par jour; il est possible qu'une capacité supérieure à celle des installations partagées soit nécessaire) : 10 000 \$ par an.

Licences de logiciels

- Même logiciel de stockage, d'extraction et d'indexage que la FTC : 230 000 \$ pour la première année et 20 000 \$ par an pour les années ultérieures.
- Logiciels supplémentaires (antivirus, systèmes d'exploitation, etc.) : 5 000 \$ par an.

Coûts de développement des logiciels

- Consultation professionnelle sur l'interface Web-utilisateur : 10 000 \$.
- Autres travaux de développement offerts par les organismes parrains.

Coûts de fonctionnement

- Dotation globale en personnel: une (1) personne (dotation combinée pour la gestion et le fonctionnement du système) : 100 000 \$ par an selon le taux de rémunération de la main-d'œuvre spécialisée.
- Autre dotation en personnel assurée gracieusement par les organismes parrains selon les besoins.
- Élaboration des politiques, planification et supervision du fonctionnement assurées par un comité d'examen non rémunéré.
- Frais accessoires (par exemple, droits d'accès) : 5 000 \$ par an.
- Sensibilisation du public et marketing assurés par les organismes parrains.

Budget (approximatif)

- Achat du matériel : 426 000 \$ pour la première année et 158 000 \$ par an pour les années ultérieures.
- Location du matériel : 470 000 \$ pour la première année et 230 000 \$ par an pour les années ultérieures.

Il serait possible de réduire les coûts de 10 000 \$ dans les deux cas si aucune capacité supplémentaire n'était nécessaire en matière d'accès au réseau.