Request for Comments : 528
Groupe de travail Réseau
NIC : 17164

J. McQuillan, BBN-NET
20 juin 1973
Traduction Claude Brière de L'Isle

Somme de contrôle logicielle dans l'IMP et fiabilité du réseau

Avec le développement du réseau ARPA de ces dernières années, et la croissance de notre expérience du fonctionnement du sous-réseau IMP (*Interface Message Processor,* processeur de message d'interface) la question de la fiabilité a pris une plus grande importance et une plus grande complexité. La présente note décrit quelques modifications qui ont été faites récemment aux programmes IMP et TIP à cet égard. Ces changements sont mécaniquement mineurs et n'affectent pas du tout le fonctionnement des hôtes, mais il valent d'être notés du point de vue logiciel, et c'est la raison pour laquelle nous avons expliqué le fonctionnement des nouveaux programmes IMP et TIP avec quelques détails. Les personnels qui gèrent des hôtes sont invités à noter en particulier les modifications décrites aux sections 4 et 5, car ils pourraient souhaiter changer leurs propres programmes ou procédures de fonctionnement.

1. Une nouvelle vision de la fiabilité du réseau

Notre vision du réseau a évolué avec la croissance du réseau lui-même. Au départ, on pensait que les seuls composants de la conception du réseau qui soient enclins à l'erreur étaient les circuits de communications, et les interfaces de modems dans les IMP sont équipés d'une somme de contrôle de CRC pour détecter "presque toutes" ces erreurs. Le reste du système, y compris les interfaces d'hôte, les processeurs d'IMP, les mémoires et les interfaces étaient tous considérés comme libres d'erreur. Nous avons dû réévaluer cette position à la lumière de notre expérience. En faisant fonctionner le réseau, nous sommes confrontés au problème d'avoir à effectuer des diagnostics à distance sur des défaillances qui ne peuvent être facilement classées ou comprises. Parmi les exemples de tels problèmes, il y a les rapports de gestion des hôtes de RFNM perdus et les messages perdus d'allocation de protocole d'hôte à hôte, les comportements inexplicables de nature transitoire dans l'IMP, et finalement le problème des défaillances – la défaillance totale d'un IMP, qui peut affecter les IMP adjacents. Ces circonstances sont peu fréquentes et sont donc difficiles à corréler avec d'autres défaillances ou avec des tentatives de remède particulières. Bien sûr, il est souvent impossible de distinguer une défaillance de logiciel d'une défaillance de matériel.

En tentant l'autopsie des défaillances, nous avons parfois trouvé que le programme de l'IMP avait des instructions incorrectes – parfois juste un ou deux bits en trop ou perdus. Il est clair qu'on peut imputer aux erreurs de mémoire presque toutes les défaillances, non seulement les pannes de programme, mais aussi les erreurs de données qui peuvent conduire à de nombreux autres syndromes. Par exemple, si l'adresse d'un message est changée dans le transit, un hôte peut alors penser que le message a été perdu, et un autre hôte peut recevoir un message en trop. Les erreurs de cette sorte entrent dans deux classes générales : les erreurs de messages d'hôte, qu'elles soient d'informations de contrôle ou de données, et les erreurs dans les messages inter-IMP, principalement les messages de mise à jour d'acheminement. Dans le courant des dernières années, il est devenu de plus en plus clair que de telles erreurs surviennent, bien qu'il soit difficile de spéculer sur où, pourquoi, et à quelle fréquence.

Un des premiers problèmes de cette sorte a été découvert en 1971. L'IMP d'Harvard avait parfois des défaillances pour une cause inconnue qui affectaient tous les autres IMP. Il a finalement été déterminé que sa mémoire était corrompue et que parfois les messages d'acheminement lus dans la mémoire par les interfaces de modem de sortie étaient tous à zéro. Les IMP adjacents interprétaient de tels massages erronés comme déclarant que l'IMP d'Harvard avait un délai de zéro pour toutes les destinations – qu'il était le meilleur chemin pour toutes les destinations ! Une fois ces informations propagées aux autres IMP, la totalité du réseau était dans la pagaille. La solution à ce problème a été de générer une somme de contrôle logicielle pour chaque message d'acheminement avant qu'il soit envoyé d'un IMP, et de la vérifier à réception par l'autre IMP. Cette somme de contrôle logicielle, en plus de la somme de contrôle matérielle du circuit, vérifie les interfaces de modem et les mémoires à chaque IMP, et protège les IMP contre toute information d'acheminement erronée. La redondance du calcul de ces somme de contrôle n'est pas très grosse car les messages ne sont échangés que tous les 2/3 de seconde.

Dans les premier mois de 1973, on a commencé à avoir de gros ennuis avec la fiabilité de certains IMP, en particulier ceux de la région de Washington. Les procédures normales d'appel et les travaux sur le terrain avec les ingénieurs de Honeywell n'avaient pas éclairci plusieurs de ces défaillances persistantes, et il a été estimé qu'il était nécessaire d'augmenter l'implication du BBN pour identifier les causes exactes de ces problèmes. Donc, durant la plus grande partie de février et mars, il y avait un ou plusieurs membres du personnel dans divers sites du réseau où des problèmes de matériel étaient suspectés. La première chose que nous avons trouvée était que le programme de fonctionnement IMP ne donnait pas suffisamment d'informations de diagnostic sur les défaillances lorsqu'elles survenaient, et que les programmes d'essai disponibles ne détectaient pas les erreurs suffisamment fréquemment pour justifier leur utilisation. C'est à dire que les

erreurs apparaissaient à une fréquence assez faible, d'une fois toutes les heures à une fois sur plusieurs jours, comparées à des taux de message d'une fois par seconde ou plus. Nous avons donc décidé d'essayer de faire tourner le programme de fonctionnement d'IMP quand il pouvait, et de rapporter plus d'informations sur les erreurs de matériel détectées, plutôt que de garder les IMP défaillants hors réseau pendant des jours.

Les modifications au programme IMP avaient deux objectifs indépendants : nous voulions rendre le logiciel moins vulnérable aux défaillances de matériel, et nous voulions que le logiciel isole les défaillances et en fasse rapport au NCC *(Network Control Centre,* Centre de contrôle du réseau (Internet)). La technique que nous avons choisi d'utiliser était de générer une somme de contrôle logicielle sur tous les paquets à mesure de leur envoi sur une ligne. Nous suspections que les défaillances de matériel dans la région de Washington survenaient entre les IMP, c'est-à-dire que les paquets étaient corrects avant leur envoi. Et donc, une somme de contrôle logicielle de mémoire à mémoire, similaire à la technique installée deux ans auparavant pour les seuls messages d'acheminement, devrait être capable de détecter ces erreurs. Le 13 mars, une nouvelle version du programme IMP a été livrée avec un code de somme de contrôle logicielle. Dans ce programme, lorsqu'un un paquet est trouvé avec une somme de contrôle incorrecte, il est éliminé et une copie des données est envoyée au NCC. L'IMP précédent retransmet le paquet, car il n'est pas retourné d'accusé de réception.

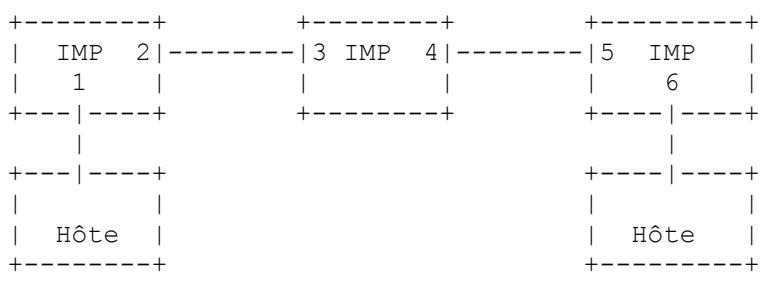
Une liste partielle des problèmes de matériel qui ont été découverts par les sommes de contrôle logicielles, et ultérieurement réglés, comporte :

* une interface de modem de l'IMP d'Aberdeen perdait plusieurs bits de plusieurs mots successifs en transférant des données en mémoire ;
* une interface de modem de l'IMP de Belvoir rajoutait un ou deux bits à chaque mot en transférant les données en mémoire ;
* une interface de modem du TIP de l'ETAC perdait le premier mot en transférant les données en mémoire ;
* une région de la mémoire de l'IMP de l'Utah changeait les deux bits de moindre poids de certains mots de façon aléatoire.

Chacun de ces problèmes résultait en deux ou trois erreurs détectées par jour. Il y avait d'autres problèmes qui n'ont pas été détectés par la somme de contrôle logicielle, telles que des interruptions intempestives. Cet ensemble de problèmes peut être expliqué par l'électronique du DMC à haut débit des IMP 316. Les trois premières machines citées ci-dessus sont des IMP 316 avec 3 interfaces de modem, et ce sont les seules machines de ce type sur le réseau. La troisième interface est dans une armoire séparée et la longueur totale du bus semble être trop longue pour l'électronique dans la conception originale. Nous investiguons à présent diverses façons de régler ces problèmes, et avons déjà obtenu quelques succès.

2. Une somme de contrôle logicielle de bout en bout sur les paquets

Cette dernière expérience, et la précédente somme de contrôle sur les messages d'acheminement, ont prouvé la valeur d'une somme de contrôle logicielle sur toutes les transmissions inter-IMP. Nous avons décidé d'étendre la somme de contrôle pour détecter aussi les défaillances intra-IMP, et de faire des sommes de contrôle logicielles sur toutes les transmissions réseau une caractéristique permanente du système IMP. Nous pouvons obtenir une somme de contrôle logicielle de bout en bout sur les paquets, sans aucun trou temporel, de la façon suivante :



* Une somme de contrôle est calculée à l'IMP de source pour chaque paquet lorsqu'il est reçu de l'hôte source (interface 1)
* La somme de contrôle est vérifiée à chaque IMP intermédiaire lorsqu'il est reçu sur le circuit provenant de l'IMP précédent (interfaces 3 et 5)
* Si la somme de contrôle est erronée, le paquet est éliminé, et l'IMP précédent retransmet le paquet lorsqu'il ne reçoit pas d'accusé de réception (interface 2 et 4)
* L'IMP précédent ne vérifie pas la somme de contrôle avant la transmission originale, pour déduire de moitié le nombre de vérifications. Mais quand il doit retransmettre un paquet, il vérifie alors la somme de contrôle. Si il trouve une erreur, il a détecté une défaillance intra-IMP, et le paquet est perdu. Sinon, la première transmission a été perdue du fait d'une défaillance inter-IMP, d'une erreur de circuit, ou a simplement été refusée par l'IMP adjacent. L'IMP précédent garde

une bonne copie du paquet, qu'il retransmet alors (interface 2 et 4).

* Après que le paquet a traversé avec succès plusieurs IMP intermédiaires, il arrive à l'IMP de destination. La somme de contrôle est vérifiée juste avant que le paquet soit envoyé à l'hôte (interface 6).

Cette technique permet d'avoir une somme de contrôle provenant de l'IMP de source à l'IMP de destination sur chaque paquet, sans lacunes dans le temps lorsque le paquet n'est pas vérifié. Toutes les erreurs sont entièrement rapportées au NCC, avec une copie du paquet en question. Cette méthode répond aux deux exigences mentionnées ci-dessus : elle rend les IMP plus fiables et tolérants à la faute, et elle fournit un maximum d'informations de diagnostic à utiliser pour l'isolation des fautes. Cette logique de somme de contrôle étendue a été installée dans le réseau le 19 juin.

Une des questions majeures sur de telles approches est leur efficacité. Nous avons été capables d'inclure la somme de contrôle logicielle sur tous les paquets sans beaucoup augmenter la redondance de traitement dans l'IMP. La méthode décrite ci-dessus implique un calcul de somme de contrôle à chaque IMP à travers lequel voyage un paquet. Nous avons développé une technique de somme de contrôle très rapide, qui ne prend que 2 ms par mot. Le programme calcule le nombre de mots dans un paquet et saute à l'entrée appropriée dans une chaîne d'instructions d'ajout. Cela produit une simple somme des mots du paquet, à laquelle le nombre de mots dans le paquet est ajoutée pour détecter les mots manquants ou mes mots supplémentaires de zéro. Avec l'inclusion de ce code, la bande passante effective de processeur d'un IMP 516 est réduite d'un huitième pour des paquets en diffusion différée de pleine longueur, d'un mégabit par seconde à 875 kilobits par seconde. C'est à dire que l'IMP a maintenant la capacité de traitement de connexion de 17 lignes en full duplex à 50 kilobits par seconde, à comparer à 20 de ces lignes sans le programme de somme de contrôle. Nous sommes conscients que cet ajout de somme de contrôle n'est pas très bonne en termes de capacités de détection d'erreurs, mais c'est tout ce que l'IMP peut supporter en logiciel. De plus, nous soulignons que le principal objectif de cette modification est d'assister les diagnostics à distance de défaillances matérielles intermittentes.

3. Somme de contrôle pour améliorer la fiabilité de l'acheminement

Nous avons mentionné précédemment les effets catastrophiques qui résultent pour le réseau tout entier d'un seul IMP qui commence à propager des informations d'acheminement incorrectes. L'expérience décrite ci-dessus impliquait une défaillance spécifique de mémoire qui ne s'est pas reproduite dans les deux dernières années, mais on comprend facilement que le problème est de nature générale. En fait, nous avons eu récemment une autre défaillance à l'échelle du réseau qui a été identifiée comme une erreur de matériel qui résultait d'un message d'acheminement erroné, après que nous ayons installé une somme de contrôle logicielle dans toutes les transmissions inter-IMP. Le problème que nous avons eu était dû à une seule instruction lésée dans la partie du programme IMP qui construit le message d'acheminement. Il en résultait que les messages d'acheminement provenant de cet IMP étaient des données aléatoires, et les IMP voisins interprétaient ces messages comme des informations de mise à jour d'acheminement. Lorsque cela arrivait, les flux de trafic à travers le réseau étaient complètement perturbés et aucun travail utile ne pouvait être fait jusqu'à l'arrêt de l'IMP défaillant.

Cette sorte de problème, l'introduction d'informations d'acheminement incorrectes dans le réseau, peut arriver de trois façons :

* Le message d'acheminement est changé dans la transmission. La somme de contrôle inter-IMP devrait capturer cet événement. Les mauvais messages d'acheminement que nous voyons dans le réseau ont de bonnes sommes de contrôle.
* Le message d'acheminement est changé lors de sa construction, par une mémoire ou une défaillance de processeur, ou avant sa transmission. C'est ce que nous appelons ci-dessus une défaillance intra-IMP.
* Le programme d'acheminement est incorrect pour des raisons de matériel ou de logiciel.

Nous avons essayé de résoudre les deux dernières sortes de problèmes en étendant le concept de somme de contrôle logicielle. Le programme d'acheminement avait été modifié pour construire une somme de contrôle logicielle pour les message d'acheminement lors de la construction du message, juste comme si il venait d'un hôte. Il est important que cette somme de contrôle se réfère au contenu prévu pour le message d'acheminement, et non à son contenu réel. C'est à dire que le programme qui génère le message d'acheminement construit sa propre somme de contrôle logicielle d'après sa propre logique, et non en lisant ce qui a été mémorisé dans la zone du message d'acheminement, mais en ajoutant le contenu prévu pour chaque entrée au fur et à mesure qu'il la calcule. Le processus qui envoie les messages d'acheminement vérifie alors toujours la somme de contrôle avant de la transmettre. Ce schéma devrait détecter toutes les défaillances intra-IMP.

Finalement, le programme d'acheminement lui-même peut subir une somme de contrôle pour détecter tout changement dans le code. Les programmes qui copient les messages d'acheminement en réception, calcule de nouvelles tables d'acheminement, et les messages d'acheminement envoyés calculent chacun la somme de contrôle du code avant de l'exécuter. Si le programme trouve une discordance dans la somme de contrôle du programme qu'il est sur le point de faire fonctionner, il demande immédiatement une rechargement de programme d'un IMP adjacent. Ces sommes de contrôle incluent le calcul de somme de contrôle lui-même, le programme d'acheminement et toutes les constantes qui s'y rapportent. Cette modification devrait empêcher qu'une défaillance matérielle sur un IMP affecte le réseau en général en arrêtant l'IMP avant qu'il ne cause de dommage sous forme de diffusion de mauvais acheminement. Une version du

programme IMP avec cette protection supplémentaire de l'acheminement a été livrée le 22 mai.

Dans les premiers mois de 1973, il y a eu plusieurs autres efforts qui visaient à améliorer la fiabilité du réseau, en plus des sommes de contrôle dans les IMP. À la même époque où nous découvrions les défaillances inter-IMP avec les paquets de somme de contrôle logicielle, nous avons commencé à remarquer une sorte de problèmes différente avec les défaillances intra-IMP. Dans ces cas, nous étions principalement confrontés à des problèmes de mémoire, et ils affectaient souvent le programme IMP lui-même, plutôt que les paquets s'écoulant à travers l'IMP. Notre première approche de ce problème a été de construire un programme PDP-1 pour vérifier les programmes IMP et TIP en cours sur un site par rapport à l'image cœur correcte détenue au PDP-1. Le programme interrogeait l'IMP avec des messages DDT, et imprimait une liste des discordances. En utilisant ce programme, nous avons déjà trouvé des défaillances de mémoire sur un site.

4. Modifications de TIP

Les difficultés de matériel que nous avons commencé à rencontrer dans les premiers mois de 1973 ont eu deux effets sur les communications d'hôte à hôte. D'abord, les défaillances intermittentes d'interface de modem, du type vu à Belvoir, Aberdeen, et l'ETAC, signifiaient que les messages étaient occasionnellement perdus par le réseau. Cette perte est rapportée à l'hôte émetteur par le message "Transmission incomplète " généré par l'IMP source ; l'hôte doit alors décider si il retransmet ou si il prend d'autres mesures. Ensuite, l'incidence plus forte que la normale des défaillances de machine signifiait que le réseau subissait parfois une "partition" de sorte qu'il n'y avait plus de chemin entre les deux hôtes communicants. (On devra cependant noter que, contrairement au concept d'origine, deux sites sont actuellement connectés au réseau par seulement un chemin ; d'autres connexions similaires sont prévues. Pour tout site de ce type, toute défaillance le long du chemin unique sera vue comme une partition.) Comme un TIP agit comme un hôte pour ses utilisateurs, sa résilience a un effet majeur sur la satisfaction de l'usager lorsque surviennent ces types de défaillances.

Avant cette époque, le programme TIP "interrompait" la connexion de l'utilisateur si il recevait une indication de Transmission incomplète de la part du programme IMP. En mars, le programme TIP (et les programmes de plusieurs autres hôtes) ont été changés pour retransmettre les messages pour lesquels l'indication Transmission incomplète était retournée ; certains hôtes (par exemple, les MULTIC) ont fait cela depuis le début. Cette modification s'est révélée être relativement simple, et nous invitons instamment les autres hôtes à envisager la mise en œuvre d'une sorte de logiciel de récupération d'erreur. D'un autre côté, il n'a pas semblé raisonnable de continuer d'essayer de transmettre lorsque le programme reçoit une indication "Destination injoignable", car cela peut venir soit d'une partition du réseau, soit d'une défaillance au site de destination. L'utilisateur interactif est, bien sûr, libre de réessayer manuellement.

Une situation différent est celle des transferts de bandes qui impliquent des TIP avec l'option de bande magnétique. Dans ce cas, l'utilisateur voudrait débuter le processus et l'ignorer ensuite jusqu'à la fin du transfert. Les partitions de réseau, même infréquentes, peuvent survenir lorsque des transferts de bandes de plusieurs heures sont en cours. Donc nous avons fait une modification significative à l'option de bande magnétique de TIP pour inclure un mécanisme de séquençage dans le protocole de transfert de bande qui permet la récupération automatique et la continuation de la transmission après la plupart des événements transitoires de réseau. Avec ce mécanisme activé, et en supposant qu'une bande est montée de "l'autre côté", le transfert complet d'une bande est possible avec une seule commande donnée de l'un ou l'autre côté. Si la connexion disparaît à mi-transfert, le logiciel de bande magnétique TIP va essayer de rouvrir la connexion jusqu'à ce que cela réussisse puis continuera alors le transfert à partir de là où il a été interrompu. En plus de la modification de l'option de bande magnétique TIP comme spécifié ci-dessus, nous avons aussi modifié le programme TENEX qui est capable de communiquer avec l'option de bande magnétique TIP de façon qu'ils restent compatibles. Ces changements ont été installés en avril.

5. Plans futurs

Nous avons examiné plusieurs des questions de fiabilité du réseau exposées ci-dessus en connexion avec le développement du nouvel IMP modulaire à grande vitesse. Cet effort conceptuel et les expériences menées avec le système IMP actuel sont bien sûr liés, et nous avons déjà décidé de plusieurs approches à entreprendre dans la nouvelle ligne des IMP :

* L'IMP aura un générateur matériel de somme de contrôle de CRC qui retournera la somme de contrôle sur une gamme de mémoire spécifiée.

* L'IMP utilisera cette facilité pour générer et vérifier une somme de contrôle de bout en bout sur les messages. Cette somme de contrôle sera donc plus exhaustive et plus apte à la détection d'erreurs que la somme de contrôle logicielle actuelle. Elle assurera un haut degré de fiabilité aux transmissions des hôtes.

* De plus, l'IMP effectuera une vérification de somme de contrôle d'un paquet à chaque bond pour fournir des

informations de diagnostic. Cette vérification sera sur une base facultative, chaque fois que le système aura des ressources disponibles pour cette vérification.

* Le code pour le nouveau système IMP sera en lecture seule (ce qui est impraticable pour les IMP 516 et 316 actuels), et le programme se fera périodiquement une somme de contrôle sur lui-même en utilisant le générateur de CRC du matériel. Nous espérons concevoir le programme de telle sorte qu'il puisse être rechargé par segments au cas de détection d'erreurs de code, sans interruption de service.

* Finalement, nous faisons des recherches sur la structure d'une somme de contrôle facultative d'IMP-hôte/hôte-IMP pour terminer la somme de contrôle d'hôte à hôte de bout en bout. Avec un tel arrangement, l'IMP et l'hôte pourraient se mettre d'accord pour vérifier les sommes de contrôle sur les messages transférés sur les interfaces entre eux, et les mécanismes de signalisation appropriés seraient fournis pour traiter les erreurs. Avec cette technique, deux hôtes pourraient s'assurer que leurs messages ont été livrés sans erreur ou autrement ils recevraient notification d'une erreur, et pourraient alors retransmettre leur message si ils le désirent.

Des détails complémentaires sur toutes modifications à l'IMP et à l'interface IMP-hôte seront publiées le moment venu.