

Optimisation de bout-en-bout du démarrage des connexions TCP

R. Sallantin ¹ E. Chaput ² A.-L. Beylot ² C. Baudoin ¹
E. Dubois ³

¹Thales Alenia Space – Toulouse

²IRIT – Université de Toulouse

³CNES – Toulouse

March 9, 2016

Plan

- 1 Contexte
- 2 Initial Spreading
- 3 Validation analytique de l'Initial Spreading
- 4 Évaluation en environnement réel de l'Initial Spreading
- 5 Conclusion

Une solution viable pour les satellites?

Satellite et TCP : je t'aime moi non plus ...

- La dépendance au RTT affecte les performances

	RTT dépendant	non RTT dépendant
début de connexion	Slow Start	
etat stable	New Reno, Reno, ...	Cubic, Compound, ...

- Or le RTT du satellite dure plus de 0.6 s.

Une solution viable pour les satellites?

Satellite et TCP : je t'aime moi non plus ...

- La dépendance au RTT affecte les performances

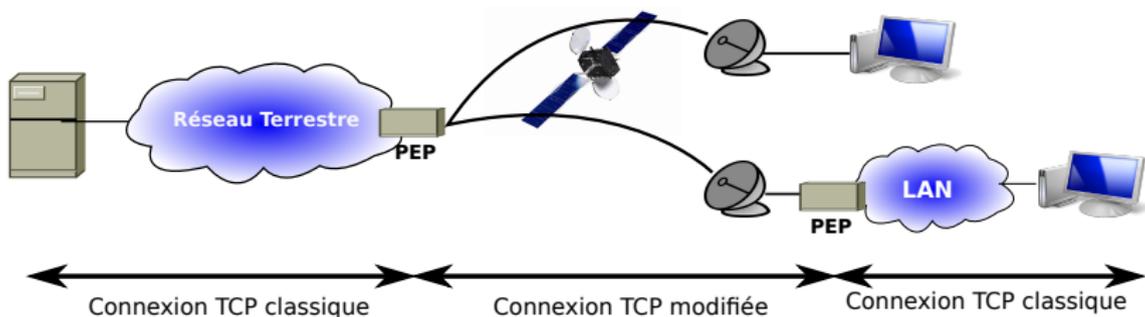
	RTT dépendant	non RTT dépendant
début de connexion	Slow Start	
etat stable	New Reno, Reno, ...	Cubic, Compound, ...

- Or le RTT du satellite dure plus de 0.6 s.

Comment s'affranchir de cette contrainte?

Les TCP-PEPs : la solution de la communauté satellite

- Objectif : Isoler et cacher le lien satellite



architecture distribuée et intégrée

Avantages et inconvénients des PEPs

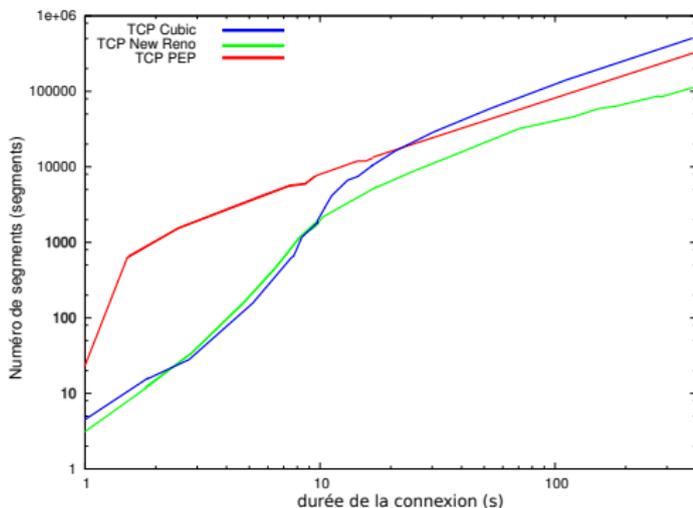
Avantages :

- Améliore la performance :
 - Réduit le RTT ressenti
 - Optimise le traitement du lien satellite
- Transparent pour les utilisateurs

Inconvénients :

- Rupture du bout-en-bout :
 - Problème de sécurité
 - Problème de mobilité
- Coûts importants :
 - Maintient et actualisation
 - Blocage protocolaire

Les TCP-PEPs faces aux solutions bout-en-bout



Comparaison réelle des solutions
TCP-PEPs, TCP New Reno et TCP

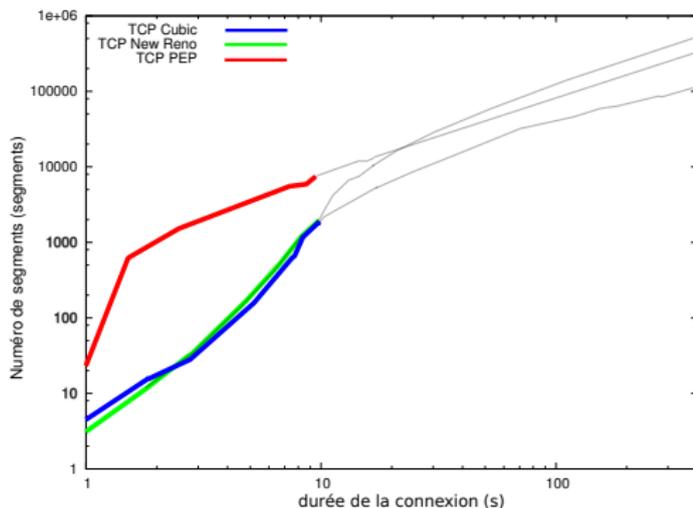
Cubic



Les TCP-PEPs faces aux solutions bout-en-bout

Connexions courtes

- Slow Start inadapté :
 - PEPs nettement plus efficace



Comparaison réelle des solutions
TCP-PEPs, TCP New Reno et TCP

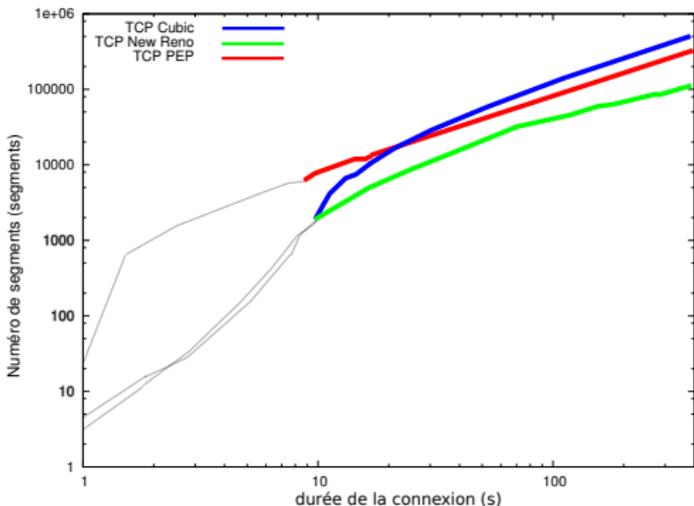
Cubic



Les TCP-PEPs faces aux solutions bout-en-bout

Connexions longues

- Nouveaux TCPs très efficaces :
 - Performance nettement supérieure à New Reno
 - Performance proche des PEPs



Comparaison réelle des solutions
TCP-PEPs, TCP New Reno et TCP

Cubic



Problématique

- TCP-PEPs performants mais très intrusifs
- Progrès importants des solutions de bout en bout pour les connexions longues
- Néanmoins, 90 % des connexions sont des connexions courtes (≤ 10 segments)

Problématique

- TCP-PEPs performants mais très intrusifs
- Progrès importants des solutions de bout en bout pour les connexions longues
- Néanmoins, 90 % des connexions sont des connexions courtes (≤ 10 segments)

2 axes d'étude

- Travailler sur l'intégration des TCP-PEPs
- **Optimiser les débuts de connexion**

Plan

- 1 Contexte
 - Analyse fine de TCP
 - Des solutions pertinentes
- 2 Initial Spreading
- 3 Validation analytique de l'Initial Spreading
- 4 Évaluation en environnement réel de l'Initial Spreading
- 5 Conclusion

Augmentation de la fenêtre initiale

Une solution intuitive :

la RFC 6928 préconise l'augmentation de l'IW à 10 segments.

J. Chu, N. Dukkupati, Y. Cheng, M. Mathis, Increasing TCP's Initial Window RFC 6928

Augmentation de la fenêtre initiale

Une solution intuitive :

la RFC 6928 préconise l' augmentation de l'IW à 10 segments.

J. Chu, N. Dukkupati, Y. Cheng, M. Mathis, Increasing TCP's Initial Window RFC 6928

- Dans les réseaux sans congestion :
=> Solution la plus rapide

Augmentation de la fenêtre initiale

Une solution intuitive :

la RFC 6928 préconise l'augmentation de l'IW à 10 segments.

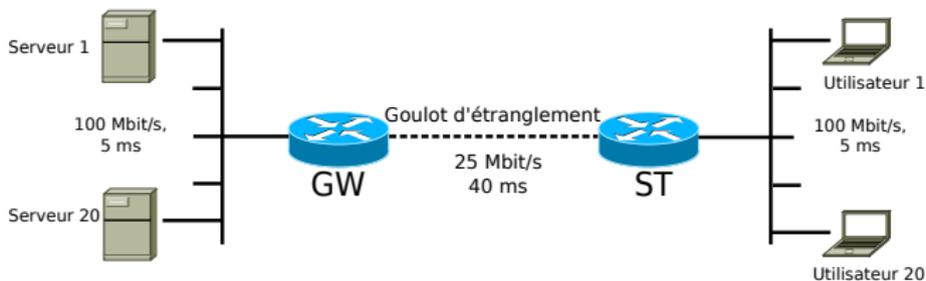
J. Chu, N. Dukkupati, Y. Cheng, M. Mathis, Increasing TCP's Initial Window RFC 6928

- Dans les réseaux sans congestion :
=> Solution la plus rapide
- Dans les réseaux avec congestion :

Quelles sont les conséquences d'une grande IW?

Topologie utilisée dans nos tests

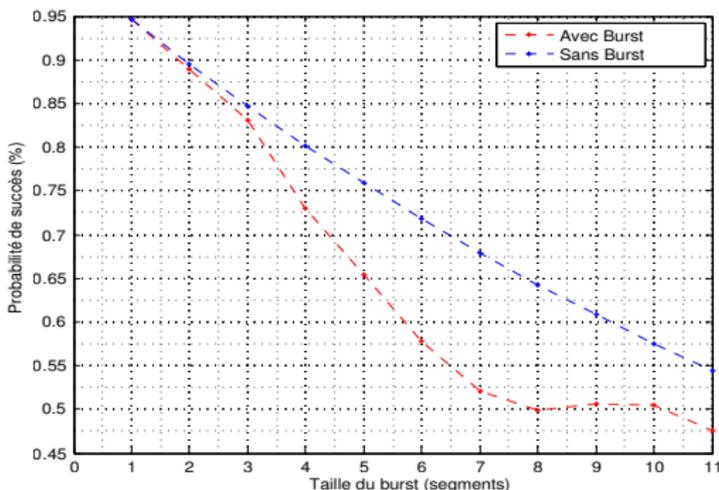
- Paramètres variables :
Débits, délais et nombre de flux en compétitions (serveurs)



Topologie "Dumbbell" paramétrée pour les tests sur les burts

conséquences des bursts de segments

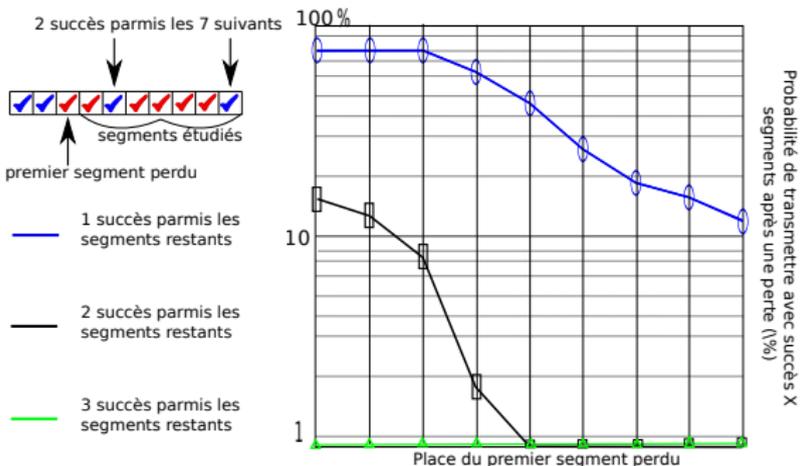
- Augmente la probabilité de perte



Probabilité de transmettre l'intégralité de l'IW en fonction de sa taille avec et sans burst

conséquences des bursts de segments

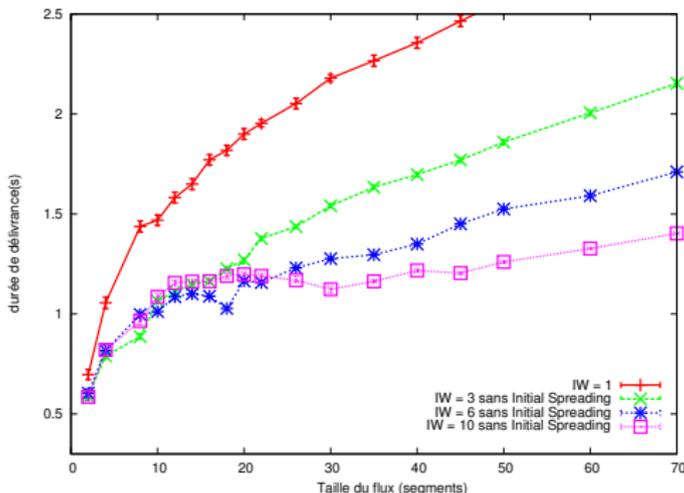
- Augmente la probabilité de perte
- Empêche les mécanismes de reprise



Probabilité d'avoir X des segments suivant la première perte reçus

conséquences des bursts de segments

- Augmente la probabilité de perte
- Empêche les mécanismes de reprise
- réduit l'intérêt d'une grande IW



conséquences de l'augmentation de l'IW dans un milieu congestionné

Pacing : la tentation inverse

Concept : Espacer la transmission de la fenêtre le long RTT

Pacing : la tentation inverse

Concept : Espacer la transmission de la fenêtre le long RTT

- Conséquences :
 - Augmentation du débit en réduisant les pertes isolées
 - Recul des premières pertes jusqu'à la saturation du réseau

Pacing : la tentation inverse

Concept : Espacer la transmission de la fenêtre le long RTT

- Conséquences :
 - Augmentation du débit en réduisant les pertes isolées
 - Recul des premières pertes jusqu'à la saturation du réseau
- Les pertes sont nécessaires au fonctionnement de TCP
- => Pacing dégrade les performances des connexions longues.

Source: A. Aggarwal, S. Savage, and T. Anderson, *Understanding the performance of TCP Pacing*,

INFOCOM 2000

Plan

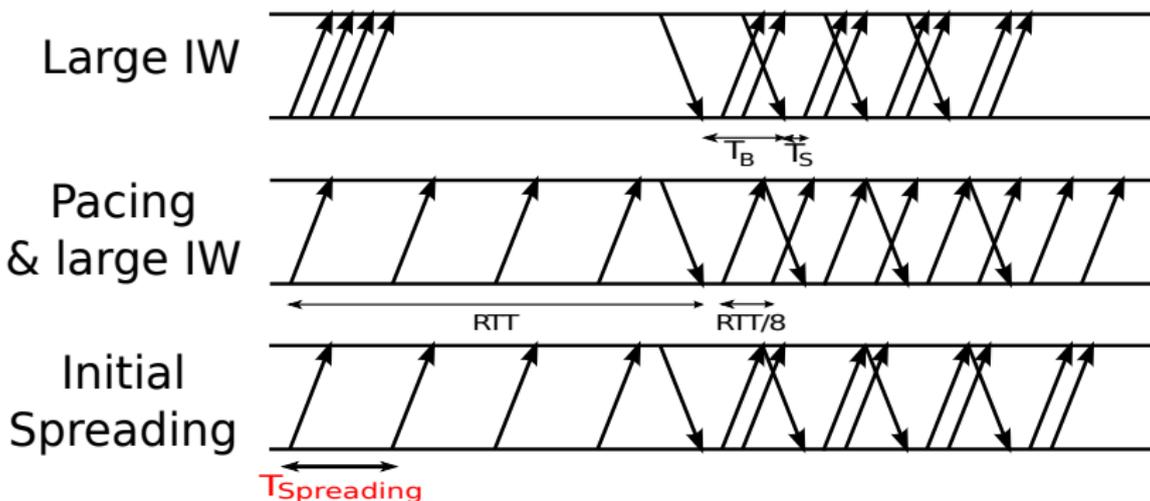
- 1 Contexte
- 2 Initial Spreading
 - présentation de l'Initial Spreading
 - Évaluation empirique par simulation
- 3 Validation analytique de l'Initial Spreading
- 4 Évaluation en environnement réel de l'Initial Spreading
- 5 Conclusion

Initial Spreading

Initial Spreading:

- Espacer l'émission de l'IW le long du premier RTT
- Laisser l'algorithme de TCP suivre son cours normal

Comportement des différents mécanismes



Émission de 12 segments avec une IW de 4 segments et 3 différents mécanismes

Nos attentes

Amélioration des connexions courtes :

- Utilisation d'une grande IW sans pâtir des bursts

Pas de dégradation des connexions longues :

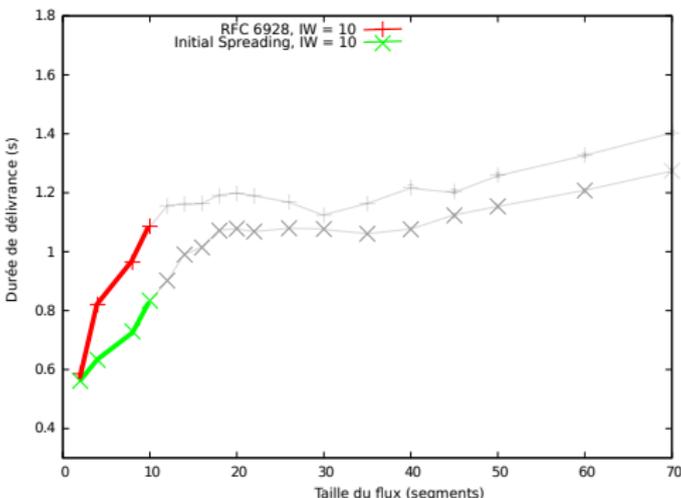
- Les mini-bursts permettent de ne pas saturer le réseau

Plan

- 1 Contexte
- 2 Initial Spreading
 - présentation de l'Initial Spreading
 - **Évaluation empirique par simulation**
- 3 Validation analytique de l'Initial Spreading
- 4 Évaluation en environnement réel de l'Initial Spreading
- 5 Conclusion

Flux courts : comportement de l'Initial Spreading

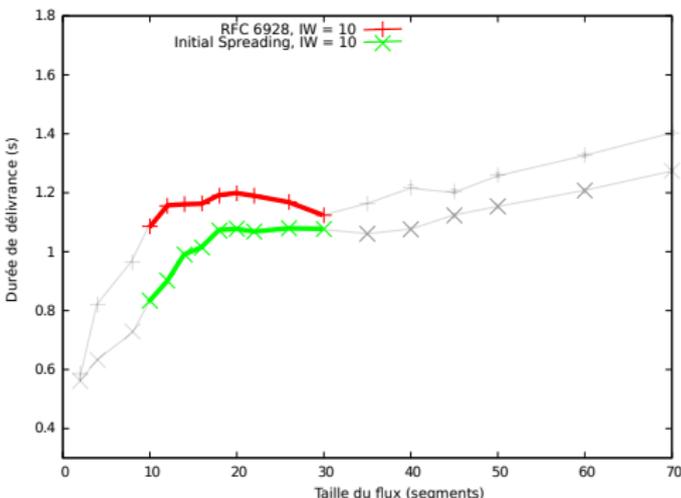
- 1 IS diminue l'impact du burst
- 2 Les mini-bursts réduisent l'écart
- 3 L'algorithme de congestion gère le débit



Analyse de l'impact des mécanismes selon la taille de la connexion

Flux courts : comportement de l'Initial Spreading

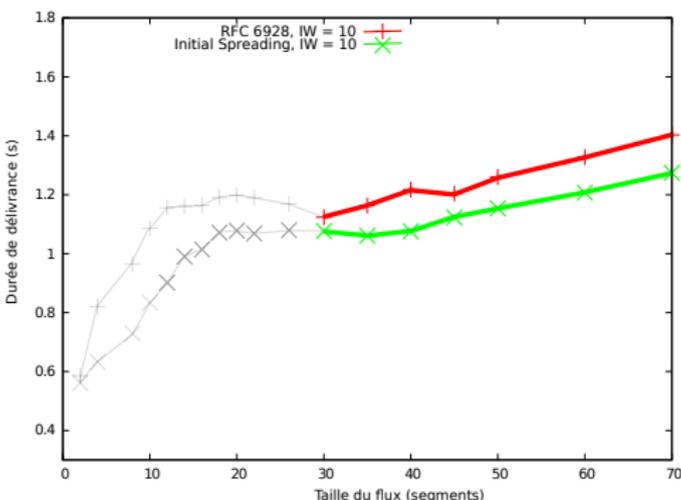
- 1 IS diminue l'impact du burst
- 2 Les mini-bursts réduisent l'écart
- 3 L'algorithme de congestion gère le débit



Analyse de l'impact des mécanismes selon la taille de la connexion

Flux courts : comportement de l'Initial Spreading

- 1 IS diminue l'impact du burst
- 2 Les mini-bursts réduisent l'écart
- 3 L'algorithme de congestion gère le débit

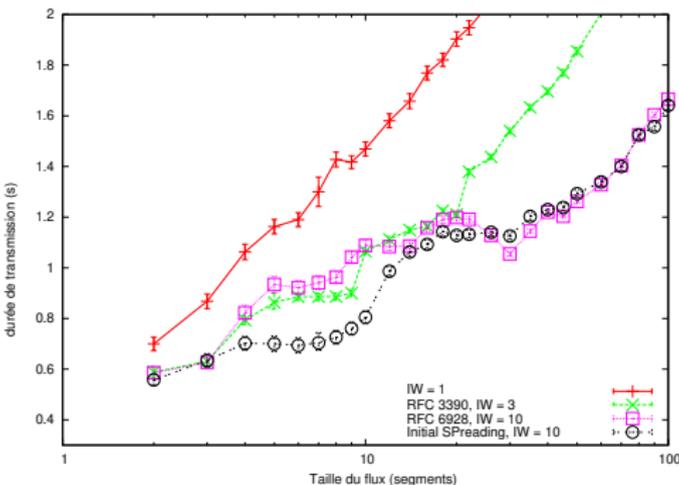


Analyse de l'impact des mécanismes selon la taille de la connexion

Performance pour les flux courts

Flux ≤ 10 segments

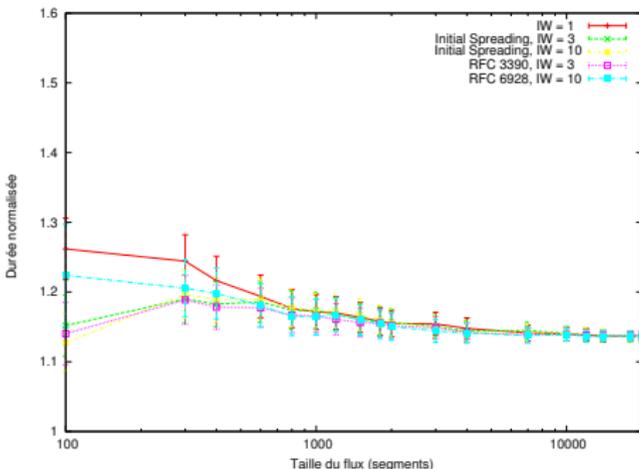
- Important gain de performance avec IS
- IW3 plus adaptée que IW10



Durée moyenne pour différentes IW avec et sans Initial Spreading

Performance pour les flux longs

- Résultats similaires avec et sans IS



Durée moyenne normalisée pour la transmission de flux longs

Plan

- 1 Contexte
- 2 Initial Spreading
- 3 **Validation analytique de l'Initial Spreading**
 - **Modèle analytique pour les connexions courtes TCP**
 - Validation par croisement du modèle et des simulations
- 4 Évaluation en environnement réel de l'Initial Spreading
- 5 Conclusion

Retour sur les bursts - Définitions

Définition classique d'un burst

L'ensemble des segments envoyés entre le premier segment d'une fenêtre et son accusé de réception

Retour sur les bursts - Définitions

Définition classique d'un burst

L'ensemble des segments envoyés entre le premier segment d'une fenêtre et son accusé de réception

=> Ne permet pas de distinguer les cas avec et sans Initial Spreading

Retour sur les bursts - Définitions

Nouvelle définition

Deux segments consécutifs d'une même connexion TCP sont considérés comme faisant partie d'un même burst si l'écart entre leur temps d'émission au niveau TCP est inférieur au temps d'émission des paquets IP correspondant au débit du lien le plus faible qu'ils traversent.

Retour sur les bursts - Types de burst

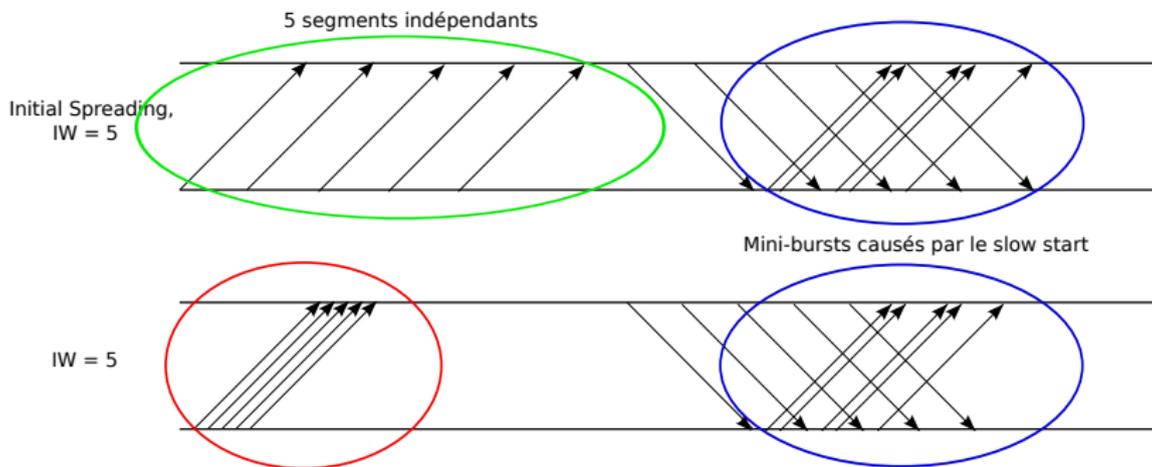


Illustration des différents bursts considérés par le modèle

Objectifs du modèle mathématique

Valider théoriquement les performances de l'Initial Spreading

- Calculer le temps moyen de transmission d'un flux courts

Un réseau supposé connu:

p : probabilité de perte par segment *indépendant*

$q = 1 - p$: probabilité de succès

R : RTT moyen

Automate à états finis - Les états

1 état est défini par :

Le nombre de segment à transmettre
La taille de la fenêtre (IW ou $CWND$)
Le type d'émission des segments

2 états initiaux:

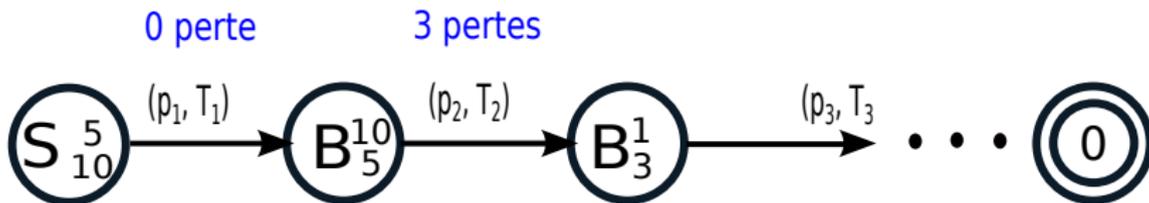
- D_i^n : i segments sont émis avec $IW = n$ sans Initial Spreading
- S_i^n : i segments sont émis avec $IW = n$ avec Initial Spreading

1 état intermédiaire:

- B_i^n : i segments sont émis avec $CWND = n$ et des mini-bursts.

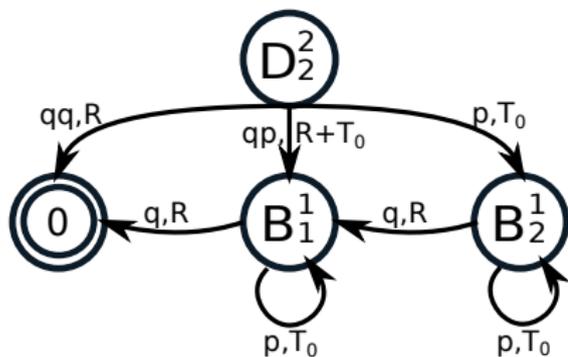
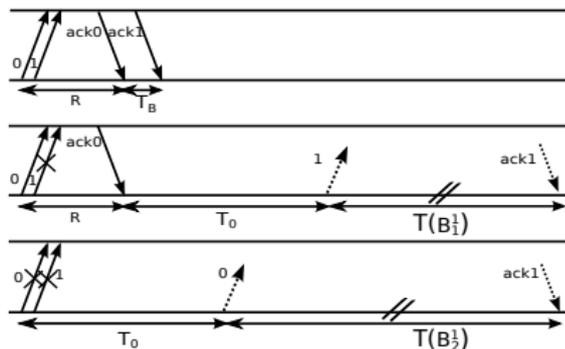
Automate à états finis - Les transitions

- 1 Tous les segments ont été acquittés
- 2 Des pertes ont été détectées



Autre représentation du même scénario

Calcul de la durée moyenne - Exemple de $\bar{T}(D_2^2)$

Automate pour D_2^2  $\bar{T}(D_2^2)$, Différents temps de transition

$$\Rightarrow \bar{T}(D_2^2) = q^2R + qp(R + T_0 + \bar{T}(B_1^1)) + p(T_0 + \bar{T}(B_2^1))$$

Durée moyenne - Flux courts

Par récursivité, nous pouvons calculer $\bar{T}(D_i^n)$, $\bar{T}(S_i^n)$ et $\bar{T}(B_i^n)$

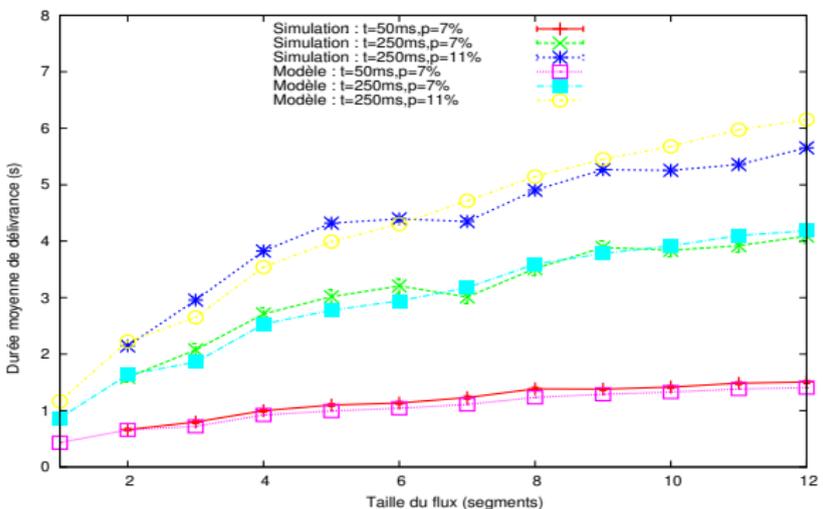
- Exemple pour $i = n$

- $$\bar{T}(D_n^n) = q^n R + \sum_{i=1}^{n-1} q^{n-i} p (R + T_0 + \bar{T}(B_i^1)) + p (T_0 + \bar{T}(B_n^1))$$
- $$\bar{T}(S_n^n) = q^n (R + (n-1) \frac{R}{n}) + p^n \bar{T}(B_n^1) + \sum_{j=1}^{n-1} q^{n-j} p^j \left\{ \binom{n}{j} (T_0 + \bar{T}(B_j^1)) + R_{i,n,j} \right\}$$

Plan

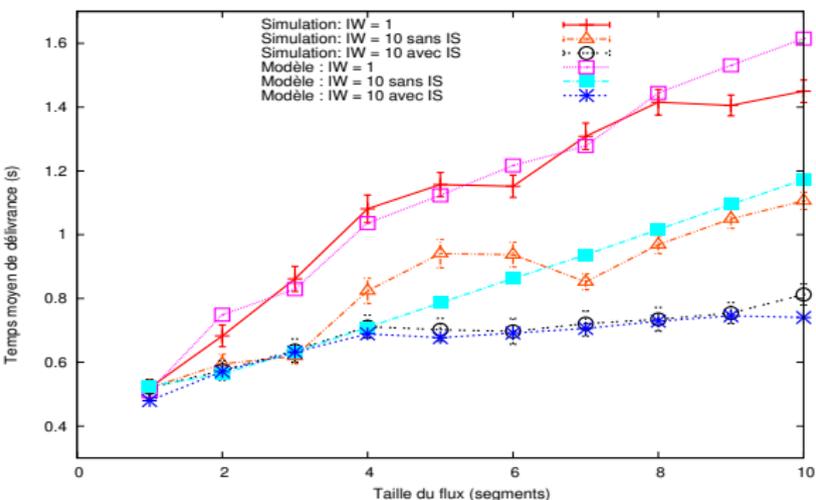
- 1 Contexte
- 2 Initial Spreading
- 3 **Validation analytique de l'Initial Spreading**
 - Modèle analytique pour les connexions courtes TCP
 - **Validation par croisement du modèle et des simulations**
- 4 Évaluation en environnement réel de l'Initial Spreading
- 5 Conclusion

Évaluation du modèle



Comparaison modèle et simulation pour flux courts avec IW=1 dans différents environnements

Validation de l'Initial Spreading



Comparaison modèle et simulation pour flux courts avec IW=1 et IW=10 avec et sans Initial Spreading

Plan

- 1 Contexte
- 2 Initial Spreading
- 3 Validation analytique de l'Initial Spreading
- 4 **Évaluation en environnement réel de l'Initial Spreading**
 - Raffinement de l'Initial Spreading
 - validation sur réseau filaire
 - validation sur réseau satellite
- 5 Conclusion

Considérations autour de l'Initial Spreading

Pour être efficace, Initial Spreading doit répondre aux contraintes suivantes :

- 1 Assurer la “non-corrélation” des segments de l'IW
- 2 Ne pas ajouter de latence inutile
- 3 Ne pas dépendre de la mesure du RTT

Considérations autour de l'Initial Spreading

Pour être efficace, Initial Spreading doit répondre aux contraintes suivantes :

- 1 Assurer la “non-corrélation” des segments de l'IW
=> $T_{Spreading}$ *suffisamment grand*
- 2 Ne pas ajouter de latence inutile
- 3 Ne pas dépendre de la mesure du RTT

Considérations autour de l'Initial Spreading

Pour être efficace, Initial Spreading doit répondre aux contraintes suivantes :

- 1 Assurer la “non-corrélation” des segments de l'IW
=> $T_{Spreading}$ *suffisamment grand*
- 2 Ne pas ajouter de latence inutile
=> $T_{Spreading}$ *le plus petit possible*
- 3 Ne pas dépendre de la mesure du RTT

Considérations autour de l'Initial Spreading

Pour être efficace, Initial Spreading doit répondre aux contraintes suivantes :

- 1 Assurer la “non-corrélation” des segments de l'IW
 $\Rightarrow T_{Spreading}$ *suffisamment grand*
- 2 Ne pas ajouter de latence inutile
 $\Rightarrow T_{Spreading}$ *le plus petit possible*
- 3 Ne pas dépendre de la mesure du RTT
 $\Rightarrow T_{Spreading}$ *borné*

T_{max} , borne supérieure de $T_{spreading}$

- 1 Segments de l'IW ne doivent pas appartenir à un burst
- 2 La latence ajoutée doit être minimisée

Débit minimal	T_{max} adapté
1 Mb/s	12ms
3 Mb/s	4ms
6 Mb/s	2ms
10 Mb/s	1,2ms
100 Mb/s	0.12ms

T_{max} , borne supérieure de $T_{spreading}$

- 1 Segments de l'IW ne doivent pas appartenir à un burst
- 2 La latence ajoutée doit être minimisée

Débit minimal	T_{max} adapté	conséquence si débit réel vaut 6 Mb/s (émission de 10 segment avec et sans IS)
1 Mb/s	12ms	ajout de 90ms
3 Mb/s	4ms	ajout de 18ms
6 Mb/s	2ms	0ms
10 Mb/s	1,2ms	création d'un burst de 4 segments
100 Mb/s	0.12ms	création d'un burst de 9 segments (RFC6928)

T_{max} , borne supérieure de $T_{spreading}$

- ① Segments de l'IW ne doivent pas appartenir à un burst
- ② La latence ajoutée doit être minimisée

Débit minimal	T_{max} adapté	conséquence si débit réel vaut 6 Mb/s (émission de 10 segment avec et sans IS)
1 Mb/s	12ms	ajout de 90ms
3 Mb/s	4ms	ajout de 18ms
6 Mb/s	2ms	0ms
10 Mb/s	1,2ms	création d'un burst de 4 segments
100 Mb/s	0.12ms	création d'un burst de 9 segments (RFC6928)

Le choix de T_{max} répond à un compromis.

algorithme final de l'Initial Spreading

- 1 Le RTT est mesuré pendant l'échange de SYN-SYN/ACK.

algorithme final de l'Initial Spreading

- 1 Le RTT est mesuré pendant l'échange de SYN-SYN/ACK.
- 2 Si n est la taille de l'IW, jusqu'à n segments peuvent être envoyés tous les $T_{Spreading}$.

algorithme final de l'Initial Spreading

- 1 Le RTT est mesuré pendant l'échange de SYN-SYN/ACK.
- 2 Si n est la taille de l'IW, jusqu'à n segments peuvent être envoyés tous les $T_{Spreading}$.

Calcul de $T_{Spreading}$

$$\text{Si } \frac{RTT}{n} \leq T_{max} \text{ Alors}$$
$$T_{Spreading} = \frac{RTT}{n}$$

Sinon

$$T_{Spreading} = T_{max}$$

algorithme final de l'Initial Spreading

- 1 Le RTT est mesuré pendant l'échange de SYN-SYN/ACK.
- 2 Si n est la taille de l'IW, jusqu'à n segments peuvent être envoyés tous les $T_{Spreading}$.

Calcul de $T_{Spreading}$

Si $\frac{RTT}{n} \leq T_{max}$ **Alors**

$$T_{Spreading} = \frac{RTT}{n}$$

Sinon

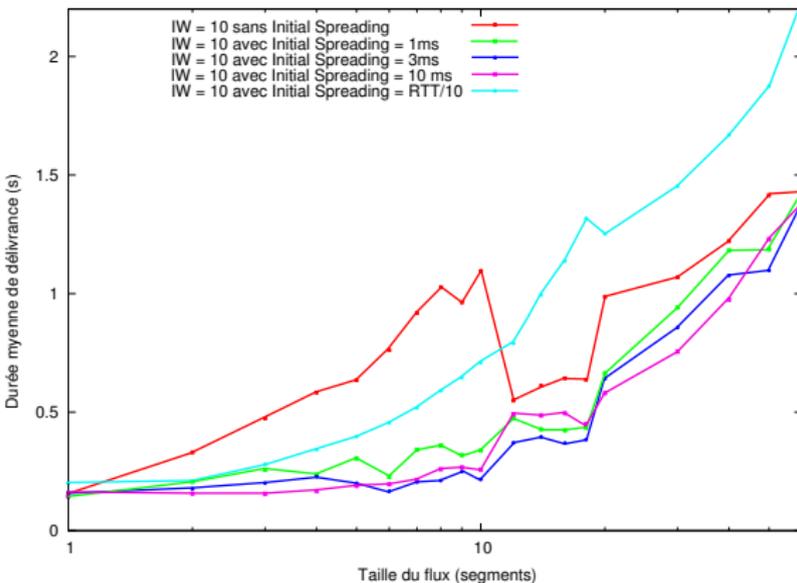
$$T_{Spreading} = T_{max}$$

- 3 Après l'émission d'IW, TCP continue de façon régulière

Plan

- 1 Contexte
- 2 Initial Spreading
- 3 Validation analytique de l'Initial Spreading
- 4 **Évaluation en environnement réel de l'Initial Spreading**
 - Raffinement de l'Initial Spreading
 - **validation sur réseau filaire**
 - validation sur réseau satellite
- 5 Conclusion

validation

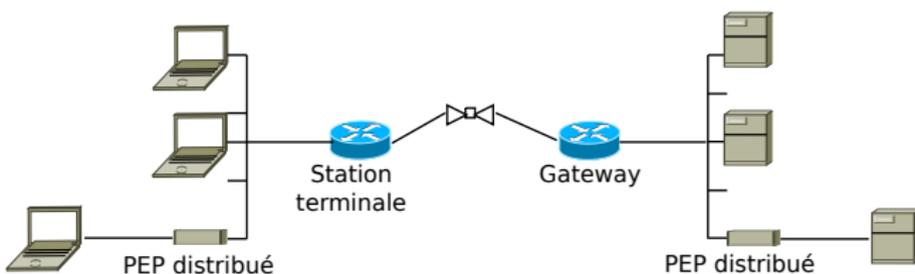


Comparaison des performances des différentes versions d'Initial Spreading et la RFC6928 sur réseau filaire

Plan

- 1 Contexte
- 2 Initial Spreading
- 3 Validation analytique de l'Initial Spreading
- 4 **Évaluation en environnement réel de l'Initial Spreading**
 - Raffinement de l'Initial Spreading
 - validation sur réseau filaire
 - **validation sur réseau satellite**
- 5 Conclusion

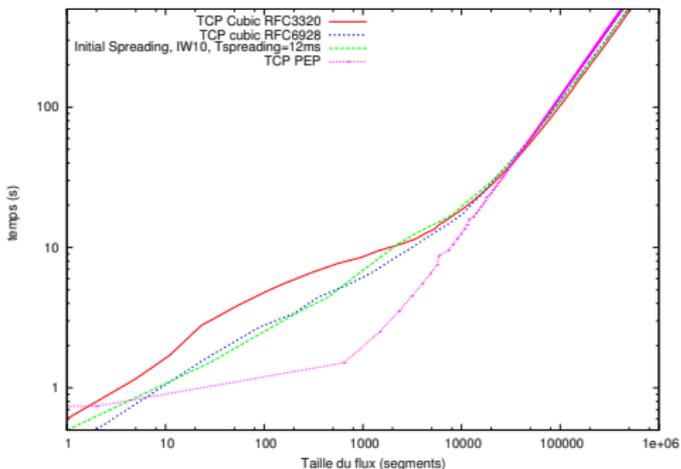
Performance de l'Initial Spreading face aux solutions bout-en-bout



Topologie pour l'émulation satellite

Initial Spreading face aux TCP-PEPs

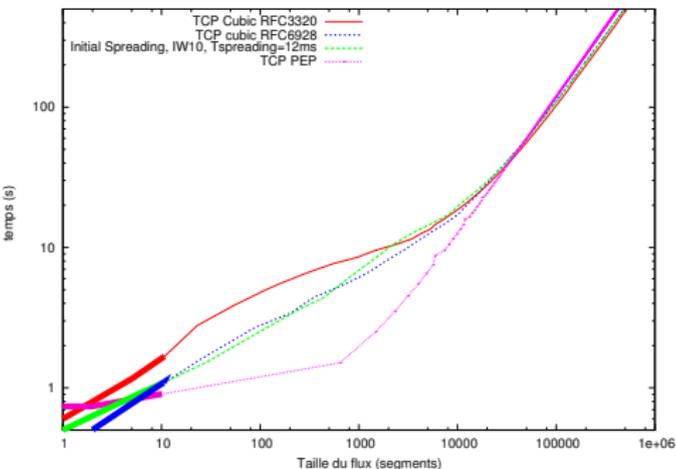
- 1 Réseau non congestionné : gains dus à l'augmentation de l'IW
- 2 Réseau congestionné : gains dus à l'IS



Comparaison des solutions existantes (RFC 3390 et 6928) avec Initial Spreading dans un réseau non congestionné

Initial Spreading face aux TCP-PEPs

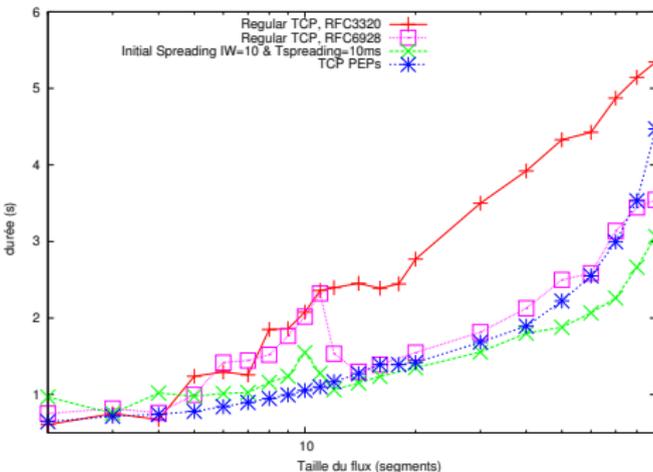
- 1 Réseau non congestionné : gains dus à l'augmentation de l'IW
- 2 Réseau congestionné : gains dus à l'IS



Comparaison des solutions existantes (RFC 3390 et 6928) avec Initial Spreading dans un réseau non congestionné

Initial Spreading face aux TCP-PEPs

- 1 Réseau non congestionné : gains dus à l'augmentation de l'IW
- 2 Réseau congestionné : gains dus à l'IS



Comparaison des solutions existantes (RFC 3390 et 6928) avec Initial Spreading dans un réseau congestionné

Plan

- 1 Contexte
- 2 Initial Spreading
- 3 Validation analytique de l'Initial Spreading
- 4 Évaluation en environnement réel de l'Initial Spreading
- 5 Conclusion
 - Conclusion
 - Perspectives

Initial Spreading

Mécanisme d'amélioration des démarrages de TCP qui rend la transmission des connexions courtes indépendante du RTT

Il permet :

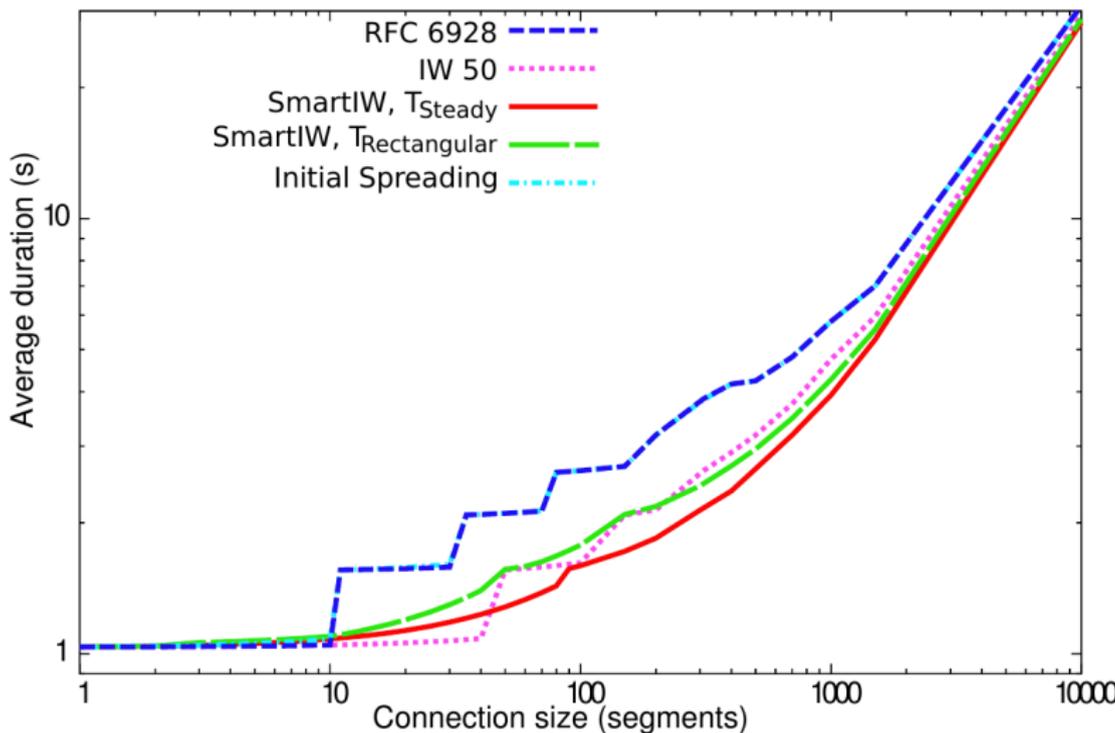
- d'envisager de se passer des solutions de type TCP-PEPs
- une réduction spectaculaire de la latence moyenne dans Internet

- Études complémentaires sur la couche MAC (cas du Wifi, ...)
- Généralisation du modèle
- Standardisation IETF
- Déploiement sur réseau satellite
- Associer notre étude des bursts aux AQM
 - au niveau des gateways
 - au niveau des routeurs

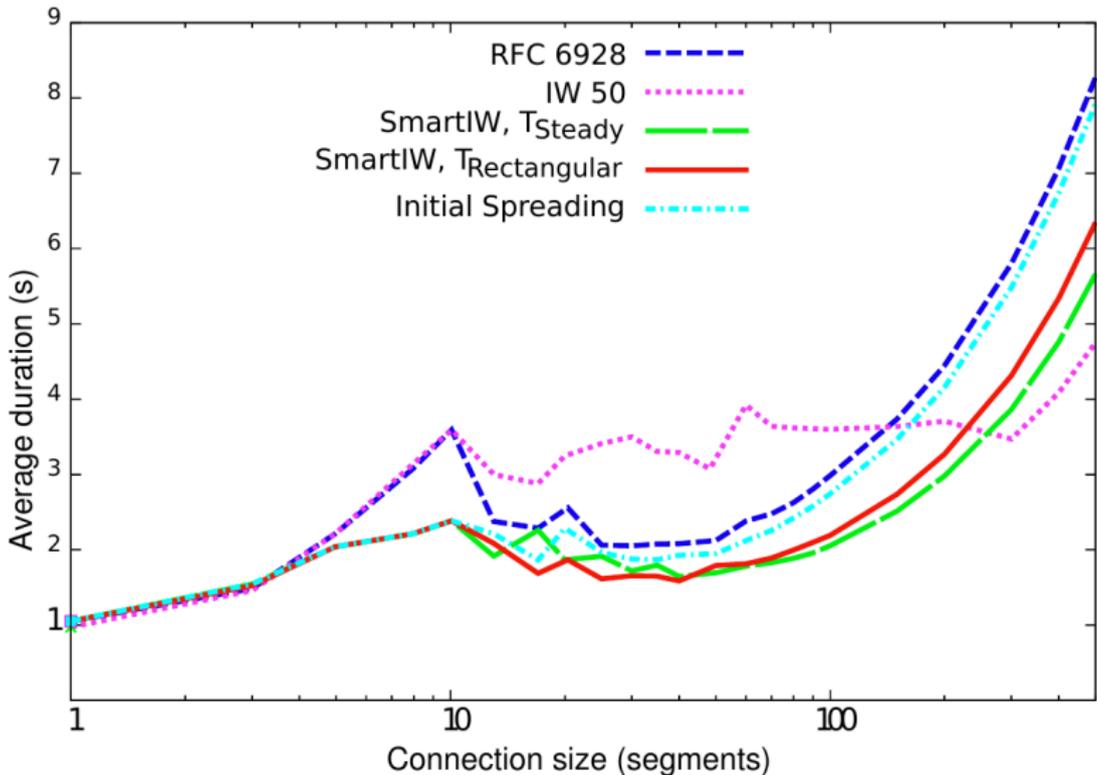
Généralisation : Smart IW

- Peu importe la taille de la fenêtre initiale
- L'important est comment on la transmet
- Utiliser le SYN/SYN-ACK pour déterminer le RTT
- Définir une fonction décrivant l'étalement des transmissions durant le premier RTT
- Cohérent avec les récentes discussions du groupe TCPM
- Voir la proposition de M. Allman

Smart IW dans un réseau satellite non congestionné



Smart IW dans un réseau terrestre congestionné



?

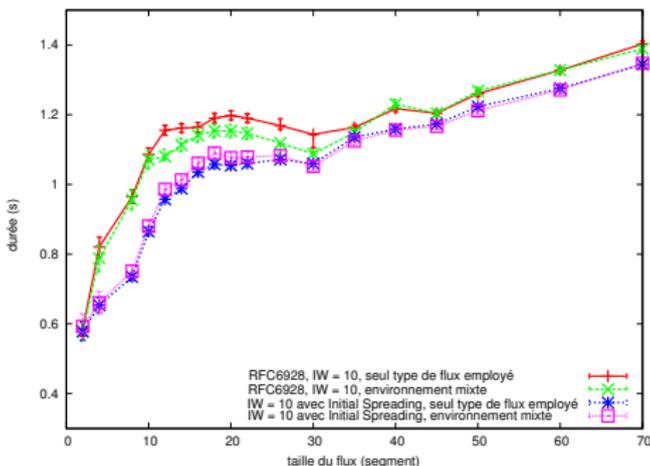
Questions ?



Équité

Contrairement au Pacing

- IS ne subit pas les autres types de flux
- IS améliore la performance globale



Constat de l'équité de l'Initial Spreading