

IETFにおけるトランスポート 関連技術の報告

慶應義塾大学SFC研究所

西田佳史

IETFの概要

- インターネット技術の標準化のために1986年から草の根的に設立された団体
 - 特別な参加資格はない。個人での参加が前提。
 - Rough consensus and running codeが意思決定のポリシー
 - MLで主に議論を進める。年3回全体ミーティングを開催する
 - インフォーマルな服装や態度が一般的
 - 率直(無遠慮とも言える)な発言がそれなりにある
- インターネットで運用される技術の標準化を行うため、Internet Protocolをベースに基本的にOSIネットワーク参照モデルで第三層以上の技術を扱う
- IETFで扱うトピックはエリアという単位で分割される。(現在7エリア)
 - Applications and Real-Time Area (art), General Area (gen), Internet Area (int), Operations and Management Area (ops), Routing Area (rtg), Security Area (sec), **Transport (tsv)**
- 各エリアは技術テーマ毎にワーキンググループ(WG)を作成し、標準化作業はそれぞれのWGで行われる。
 - WGの構成人数やトピックの大きさなどはWGによって様々
- IRTF (Internet Research Task Force)という姉妹団体がある
 - IRTFは標準化は行わないが、IETFと並行しながら長期的視点でインターネット技術の議論を行う。

発表者とIETFの関わり

- 学生時代よりトランスポートプロトコルの研究開発を継続的に行っている
- 2008年頃から積極的にIETFに参加
 - RFC6582, RFC6675, RFC6789などの著者となる
 - その他トランスポート関連のWGでI-Dを定期的に発表している
- 2008年からMultipath WGの Co-Chairを務める
- 2012年からTCPM WGの Co-Chairを務める
- 2008年から2016年まで Transport Area Directorateのメンバを務める
- 2016年からTransport Areaの Transport Area Review Team (TSVART)のメンバを務める
- 2015年からIANAの Port Request Reviewerのメンバを務める

IETFトランスポートエリア概説

- IETFの7エリアの1つ

- 4層技術を中心とした技術の開発、標準化を行う
- 主なWG
 - TCPM ... TCPの比較的小規模な拡張と問題点の修正の標準化を行う
 - TSVWG...TCPに特化しないトランスポート関連の技術やRSVPなどの標準化を行う
 - TCPINC ... TCPの暗号化拡張の開発、標準化を行う
 - Multipath TCP ... TCPのマルチパス通信拡張の開発、標準化を行う
 - QUIC ... QUICプロトコルの開発、標準化を行う
 - TAPS .. トランスポートサービスの抽象化されたAPIの開発、標準化を行う
 - DTN .. Delay Tolerant Networking 技術の開発、標準化を行う

- IETFトランスポートエリアと関連の深いIRTFのRG

- MAPRG .. プロトコルの挙動の解析やインターネットのトラフィック収集、解析に関する議論を行う
- ICCRG .. インターネットにおける輻輳制御技術に関連する議論を行う

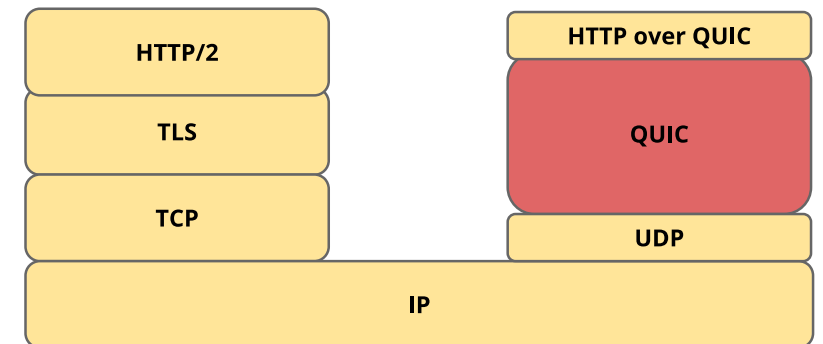
最近のトランスポート関連技術のホットトピック

- QUIC
 - TCPに代わる新しいトランスポートプロトコル
- 新しいTCP輻輳制御技術
 - 従来よりも高スループット、低遅延を実現する輻輳制御アルゴリズム
 - TCP RACK, TCP BBR
- ECN (Explicit Congestion Notification)
 - インターネットの渋滞状況をより正確に把握する技術
- マルチパス通信
 - 複数の通信経路を利用した通信

QUIC

- QUICとは

- TCPに代わる新しいトランスポートプロトコル
- HTTPでの利用を強く意識した設計
 - UDPの上位層プロトコルとなる →NAPT問題への対策
 - TCP+TLSよりも低遅延な通信を実現
 - 暗号化機能をサポートし、全ての通信で暗号化を行う
 - TCPの様々な問題点に対する対策を取り入れた設計
 - Apple, Facebook, Microsoft, Google, Mozilla, 等で実装が進んでいる
 - QUICを利用するHTTPの新バージョンは、HTTP/3として開発が進められている。(HTTP/2の開発も継続)
 - HTTP over TCPとHTTP over QUICを明確に区別するためバージョン番号をアサインした



<https://conferences.sigcomm.org/sigcomm/2017/files/program/ts-5-1-QUIC.pdf>より

QUICの標準化の状況

- QUIC WGで標準化を進めている
- 現時点の主なトピック
 - V1のリリースに向けて仕様のフリーズを図っている
 - Multipath対応などはV1には含まれない
 - QUICとmiddleboxとの関係性に関する議論
 - QUICはmiddleboxの影響を避けるためヘッダの大部分を暗号化しているが、これが問題となることもある
 - 暗号化されないものは、バージョン番号やコネクションIDくらいしかない
 - QUICコネクションをモニタしても通信の状態が全くわからない。(通信状況の把握が困難)
 - SPIN Bitの導入で対応
 - ヘッダにRTT観測用の1ビットを追加することにより、トラフィック観測からQUICセッションのRTTを計測を可能にする。
 - Version Ossification問題
 - Version 1以外のヘッダを持つQUICパケットがmiddleboxによって廃棄されてしまう問題
 - 議論は継続中

QUIC WG以外でのQUICに関する議論

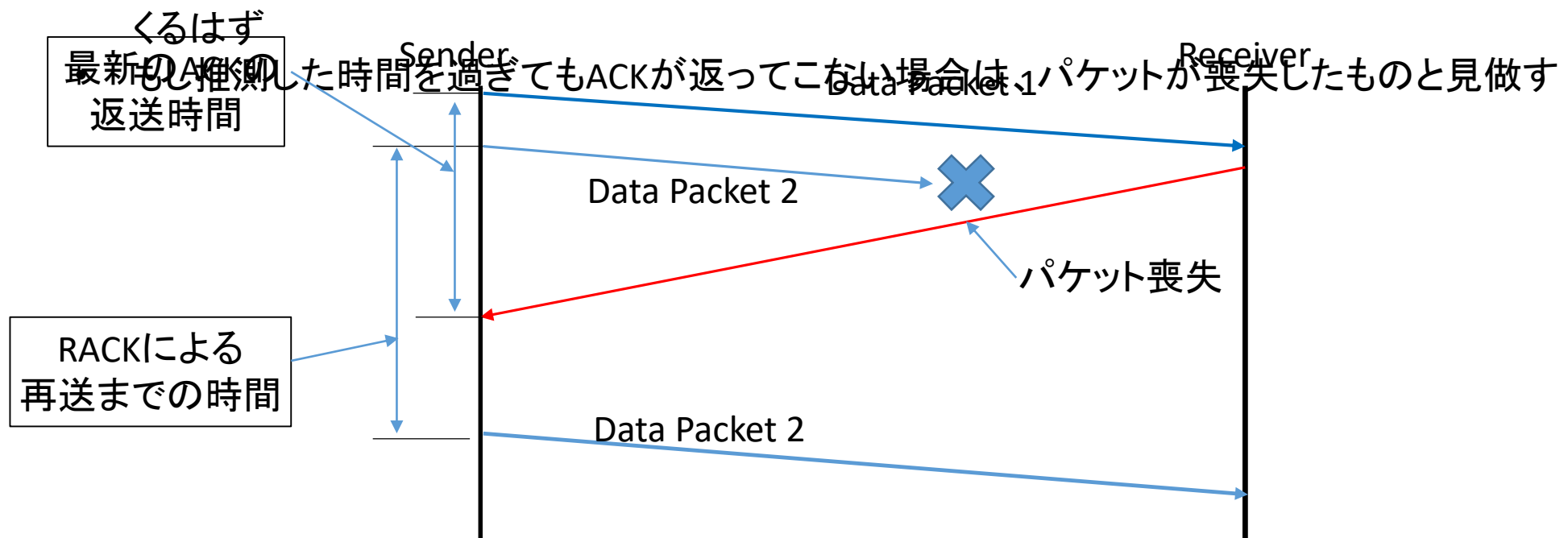
- IETF 105
 - IRTF Applied Networking Research Workshop 2019
 - A Performance Perspective on Web Optimized Protocol Stacks: TCP+TLS+HTTP/2 vs. QUIC.
 - Performance Measurements of QUIC Communications
 - QUICを使った性能解析に関する研究発表
 - MAPRG
 - Measuring QUIC Dynamics over a High Delay Path
 - 衛星通信のような高遅延ネットワークにおけるQUICの性能評価に関する発表
 - Trials and tribulations of migrating to IETF QUIC
 - QUICを実際にデプロイ、運用した経験に関する発表
- IETF 106
 - WebTransport BoF
 - WebSocketに代わる技術に関する議論
 - MAPRG
 - QUICのデプロイ状況に関する報告 (e.g. 50%のyou tubeトラフィックがQUIC等)
 - QUIC performance over high BDP BoF
 - 高遅延ネットワークにおけるQUICに関する研究開発の進め方等の議論
 - TSVAREA (トランスポートエリアの全体ミーティング)
 - QUICの今後の進め方、WGでやるべき事、やるべきでない事等に関する議論

新しいTCP輻輳制御技術

- Cubic, NewRenoなどよりも高スループット、低遅延を実現するアルゴリズムの開発
- TCP RACK (Recent ACK)
 - TCPの新しいパケット喪失検出アルゴリズム
 - 早期にパケット喪失を検出することで、素早く再送を行い、データ転送の効率をあげる狙い
 - パケット喪失の検出アルゴリズムなので従来のCubicやNewRenoに取り入れて利用できる
- TCP BBR (Bottleneck Bandwidth and Round-trip propagation time)
 - ボトルネックリンクの帯域幅とRTTの伝搬時間を推測しながら通信
 - 推測した値に応じてデータ転送速度を調節することで、輻輳を起こさない最適速度で通信を行う狙い
 - Googleが開発し、Google, Netflixなどで採用されており、注目を集めている

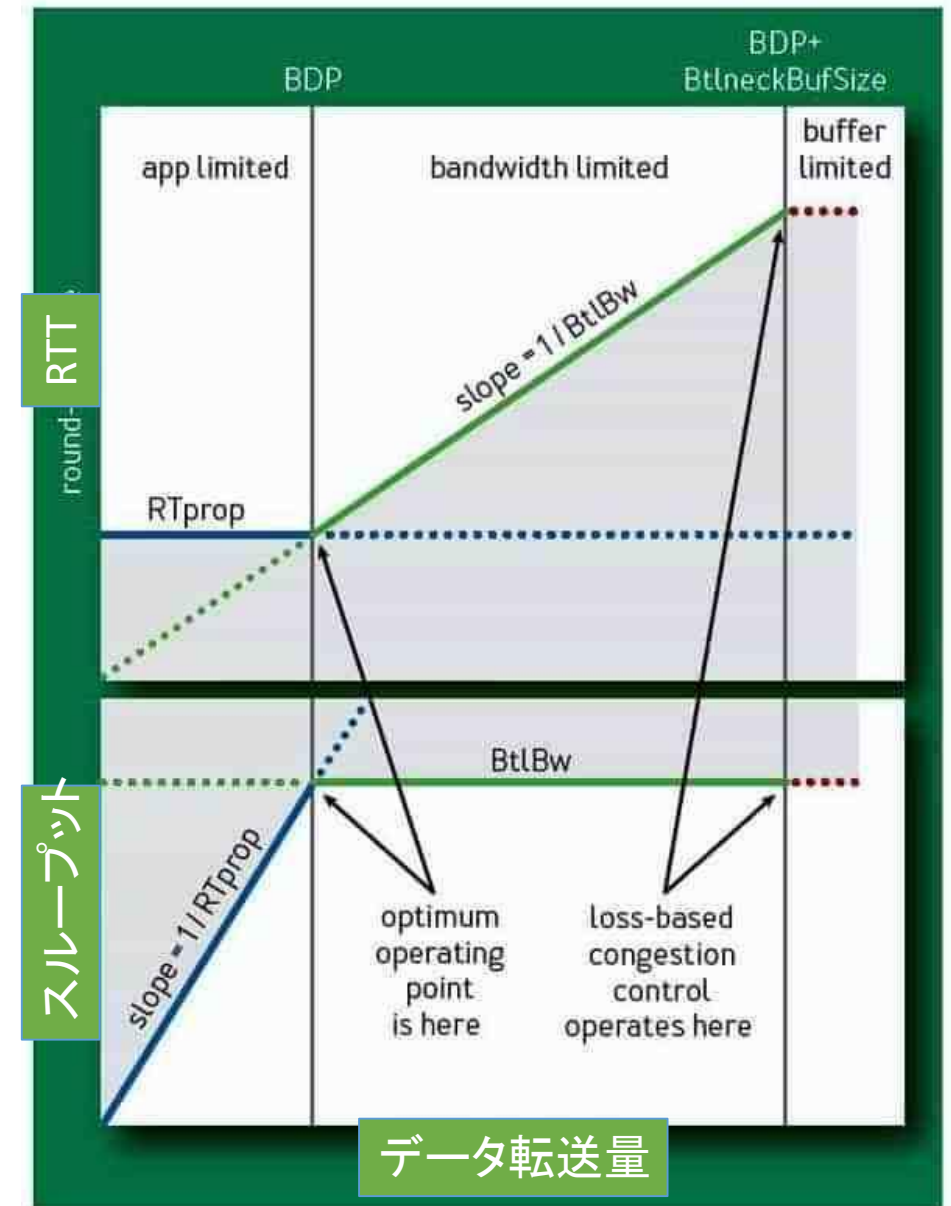
TCP RACK (Recent ACK)の概要

- パケットロスを早期に検出するアルゴリズム
 - 従来のアルゴリズムは以下の2つの方法でパケットロスを検出していたが時間がかかる
 - 再送タイムアウト ... 1秒程度の時間がかかるため検出が遅い
 - 3つ以上の重複確認応答の受信 ... 状況によっては重複確認応答が3つ以下しか返信されない
 - RACKは最新のACKの返送状況から、他のパケットの返送時間を推測
 - 最新のACKはこのくらいの時間で返ってきたので、他のパケットのACKも同程度の時間で返ってくるはず



TCP BBRの概要

- TCP BBRと既存の輻輳制御技術の違い
 - 既存の輻輳技術 ... パケットロスがあればネットワークに輻輳があると判断し、転送速度を調節
 - TCP BBR ... **パケットロスに関わりなく**、BBRが推測した可用帯域に基づいて転送速度を調節
- データ転送量がネットワークの許容量を超えた時点でスループットは頭打ちになり、RTTだけが増加する
 - ネットワークのバッファが溢れた時点でパケットロスが起きる
- BBRの目標は、スループットが頭打ちになった時点を検出すること
- V2では、パケットロス率が5%を越える場合と場合に転送速度を制御するロジックが追加された
 - ルータのバッファが少ないケースで転送性能が悪化する問題などに対処するため

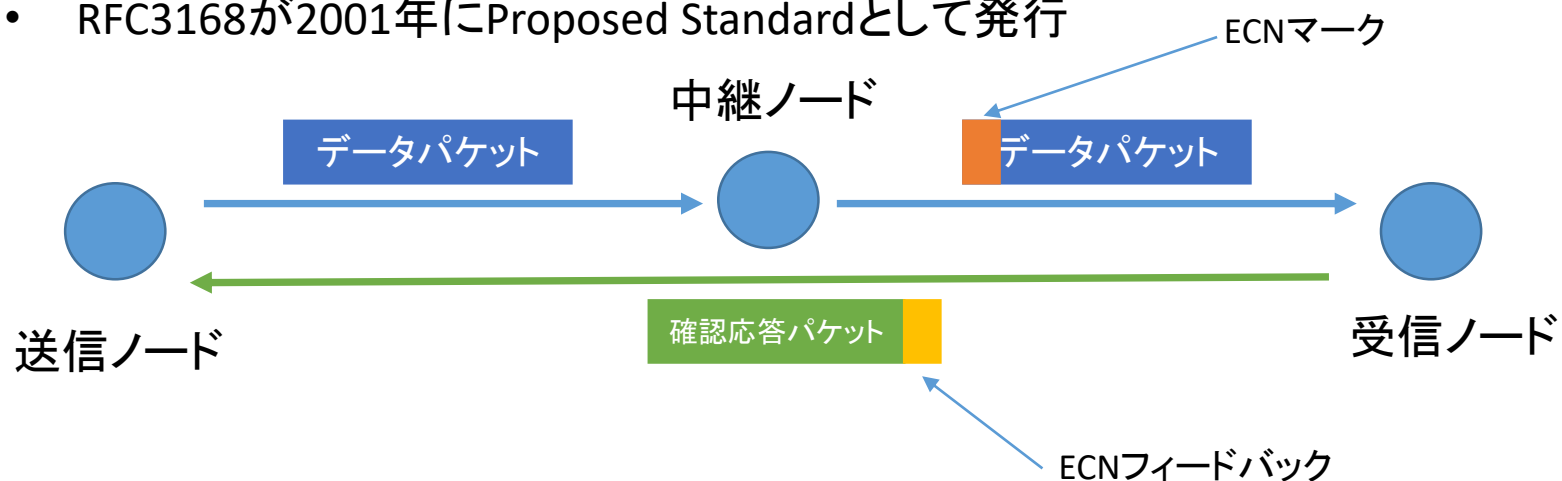


TCP RACK、BBRの標準化動向

- TCP RACK
 - TCPM WGで標準化の議論を進めている。議論はほぼ終了し、Proposed StandardとしてRFCが発行される見込み
 - Google, Microsoft, Netflixなどが強い支持を表明
- TCP BBR
 - 従来のTCPの輻輳制御アルゴリズムと大きく異なるため、現時点ではIETFでなくIRTFのICCRGなどで議論が進められている
 - 開発状況の報告や性能解析に関する議論がそれなりの頻度で行われている
 - V2の仕様をまとめてInternet-Draftとして発表するプランをGoogleの開発者が表明

ECN (Explicit Congestion Notification)

- ECNとは？
 - 中継ノードが輻輳の状況をエンドノードにフィードバックする仕組み
 - 中継ノードが自ノードのデータ送信待ちキューの状況をチェックし、ある一定以上のキュー長がある場合、ネットワークが輻輳しているものとして、データにECNマークをセットする
 - ECNのマークが付いたデータを受信したエンドノードは、ECN受信したというフィードバック情報をACKなどに格納して送信側に伝達する
 - ECNのフィードバックを受信した送信側は、転送速度を減少させ輻輳の悪化を回避する
 - パケットロスが起きる前に輻輳の前兆を検出できる
 - パケットロスを起こさずに輻輳を回避できる
 - 2000年頃に標準化作業が進められた
 - RFC3168が2001年にProposed Standardとして発行



ECNの状況

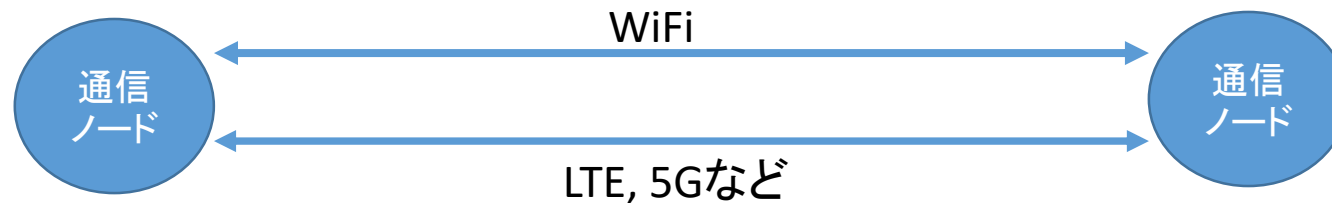
- ECNの技術は昔からあるが、それほど普及していなかった
 - ECNマークの付いたパケットを廃棄するルータが存在していた
 - ECNマークの付いたSYNパケットを無効なパケットと見做すTCP実装が存在していた
 - ECNを利用するコネクションの方がECNを利用しないコネクションよりもスループットが低いケースがあった
 - ECNを利用すると輻輳が起こる前にデータ転送速度を落とす
 - RFC3168ではECNマークとパケットロスを同等に扱う
- 2015年前後からECN再開発の機運が高まってきた
 - 近年のルータやTCP実装の多くはECNをサポートしている。Activateしていないだけ
 - Microsoftが新しいECNのアプローチであるData Center TCP(DCTCP)を発表し、関心を集めた
 - 単一のECNマークを利用するのではなく、ECNマークが付けられる頻度から輻輳の状況を推測
 - ECNマークの頻度が低い→データ転送速度を少しだけ減少させる

ECNの標準化動向

- TSVWGにおける標準化活動
 - Low Latency, Low Loss and Scalable throughput (L4S)
 - ECNマークの頻度情報を元に転送速度を調節。Accurate ECNを利用
 - Some Congestion Experience (SCE)
 - Codel(bufferbloatに適したキューイングアルゴリズム)を利用するMiddleboxを想定したECNアルゴリズム
- TCPM WGにおける標準化活動
 - Accurate ECN
 - TCPでECNマークの頻度情報をフィードバックする仕組み。L4Sでの利用を意識
 - Generalized ECN
 - TCPのSYNパケットやACKパケットにECNマークを付与する提案
 - TCP Feedback for Some Congestion Experience (SCE)
 - TCPでSCEの情報をフィードバックする仕組み
- SCEとL4Sは同時に利用できない仕組みなので、今後どのように棲み分けるかが課題
 - 通信経路上にSCEをサポートするものとL4Sをサポートするものが混在するとECNマークの情報が混乱する

マルチパス通信

- マルチパス通信とは
 - 複数の通信経路を利用したデータ転送
 - TCPなどで1つのコネクションで複数のアドレスを利用可能にする
- マルチパス通信の利点
 - 複数通信経路の利用により、高スループット通信、ロバストな通信 (WiFi + 5Gなど)
 - 複数経路を同時に使って高いスループットを得る
 - 常に最も遅延の小さな経路を選択し、低遅延通信を行う
 - IPアドレスが通信中に変化する移動通信に対応できる
 - 一つの経路に問題があっても、他の経路で通信を継続



2つの通信回線を利用するマルチパス通信の例

マルチパス通信の標準化動向

- Multipath TCP
 - マルチパス通信を可能にするTCPの拡張機能
 - 複数のTCPコネクションを1つのセッションに属するsubflowとして制御する
 - TCPオプションを用いて制御情報を通信ノード間で交換する
 - TCPの拡張なので、既存のTCPの高速化技術(TSO,LRO)などがそのまま利用可能
 - Multipath TCP WGで標準化を進めている
 - 旧バージョンの仕様は RFC6824として発行
 - 現在は新バージョンの標準化を行なっている
 - 旧バージョンから得られた知見による仕様のアップデート
 - セキュリティの強化
 - AppleはSiriなどで低遅延通信を行うために積極的にMPTCPを利用している
 - Linux実装をのLinuxの公式カーネルにマージする活動が進められている
- QUIC
 - QUIC WGでマルチパス機能のサポートが議論されている
 - V1ではサポートしないことが決定しているが、V2ではサポートされる期待が高まっている

総括

- 105,106回会合におけるトランスポート関連技術のトレンド
 - QUICは依然かなり高い関心を集めている印象
 - QUIC WGのミーティングは常に100人以上の参加者
 - TSVAREA,TSVWG,TAPSなどのWG でQUICに関する発表が増えている
 - V1は仕様がほぼ固まりつつあり、V2に向けた議論が進められている
 - BBRへの関心の高まり
 - Google以外の組織でも(Netflix, Verizon, Dropbox)性能評価、解析が進められている
 - FacebookはCOPAというdelay basedのcongestion controlとBBRの性能比較を行なっている
 - ECNの拡張仕様の標準化が進んでいる
 - ECNが普及すれば今後の輻輳制御技術に影響を与える可能性がある
 - Accurate ECNとSCEの棲み分けが課題