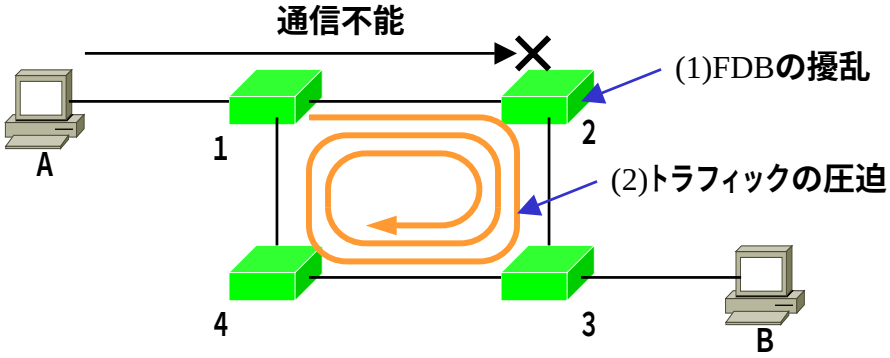
広域イーサネットはどう変わったか？
ＪＡＮＯＧ１５
２００５年1月２１日

安藤　雅人

やっぱり、イーサネットの最大の敵はループ

- ループが発生したイーサネット内部では通信がほぼ出来なくなる。
  （広域イーサネットでもこれは同じ、規模が大きい分影響も大きい）



- ループによって通信が出来なくなる２つの要因。
  (1)ループが発生すると、ＦＤＢ（Forwarding Data Base=MAC学習テーブル）の擾乱が発生する。
  → フレームが正しい宛先に転送出来なくなる。
  (2)ループ内でフレームが増殖し、帯域の圧迫が発生する。
  → 帯域不足によりフレームのロスが発生する。

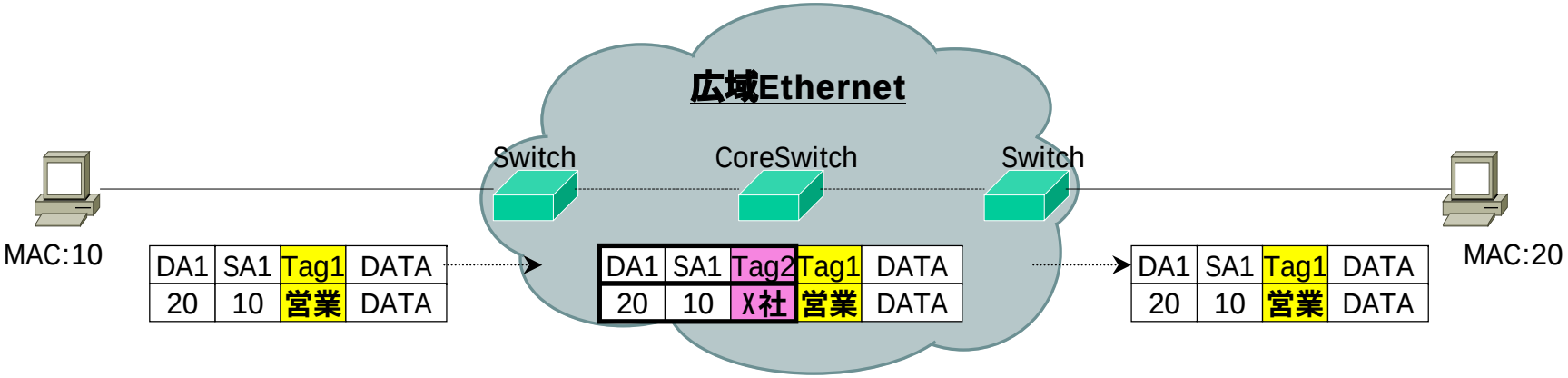
「広域イーサネットではループを排除する事が至上命題」

ループと戦う為の武器

- **ループフリーな論理トポロジーを維持する機構**
  - STPとか、XXRP、とかリングみたいなL2の冗長化機能
    （これらの機構が上手く動作しない場合に発生するループが問題）
  - ループしない物理構成
    （ルータで冗長化出来るなら、わざわざL2で冗長化しなくともよい）
- **ループの発生を検出し、可能ならば論理トポロジーに働きかける機構**
  - 隣接ノードのコントロールプレーンの正常性を疑う方法（Cisco の Root Guard, Loop Guard等）
  - 実際の転送状況をモニタする方法（ExtremeのELRP、その他ベンダのループ検出用のフレームを定期的に送信する方法）
  - ノードの故障やＣＰＵの過負荷によるループはかなり防げます、普通のEthernetでも冗長を組まれる方は、出来る範囲で是非やっておきましょう。
- **ループしたフレームを検出し、フレームを破棄する機構**
  - JANOG10で構想を発表した実装Ether over Ether(EoE)ですが、ループ防止機能も盛り込んで広域イーサ用に実装して使ってます。（パワードコムなど）
- **ループを検出する機構**
  - FDBの書き換わり検知
  - ループ検出用のフレームを定期的送信する方法

今までの広域イーサは1Qin1Qがほとんど

- 現在提供されている広域イーサではIEEE 802.1Qの技術（あるいは類似の技術）を改造し、Tagを多重化しユーザのVLAN Tagを透過するようにした１Qin1Q（VMAN)が広く使われている。

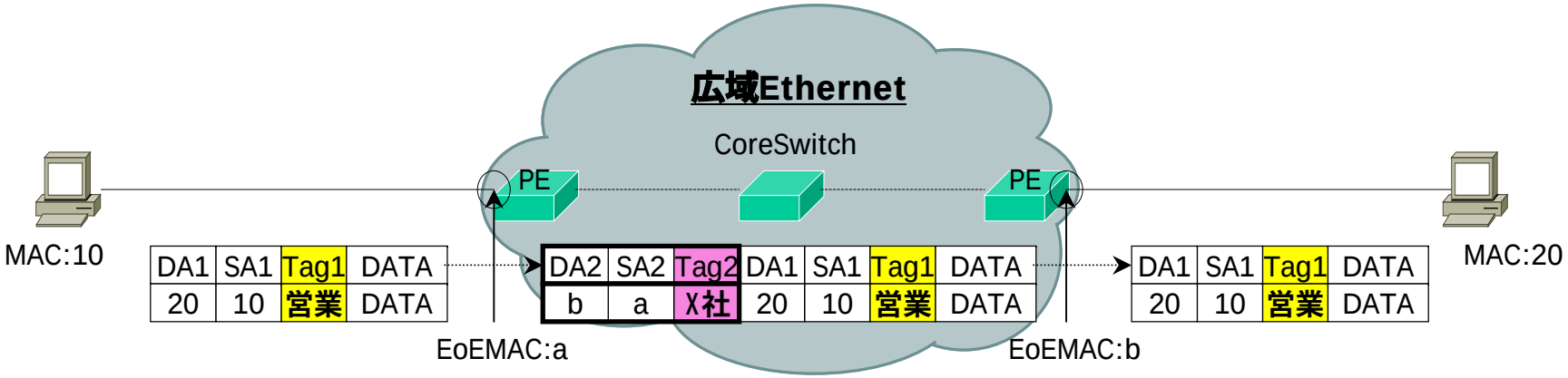


１Qin1Qのシステムにおいても、「ループフリーな論理トポロジーを維持する機構」に「ループの発生を検出し、可能ならば論理トポロジーに働きかける機構」を組み合わせて使用すればかなり、ループは防げる。

ループ対策も含めEther over Ether(EoE)を導入

802.1Q Tag VLANを使ったVLAN VPNの改良方式

- ユーザが使用するMACアドレスとは別に、EoE内の転送に使用するMACアドレスを別に準備しそのMACアドレスを持つイーサネットフレームでユーザのイーサネットフレームをカプセル化して転送する方式。２００２年７月のJANOG10で広域イーサネット用のシステムとして構想を発表、昨年より実際の広域イーサネットで利用し始めている。

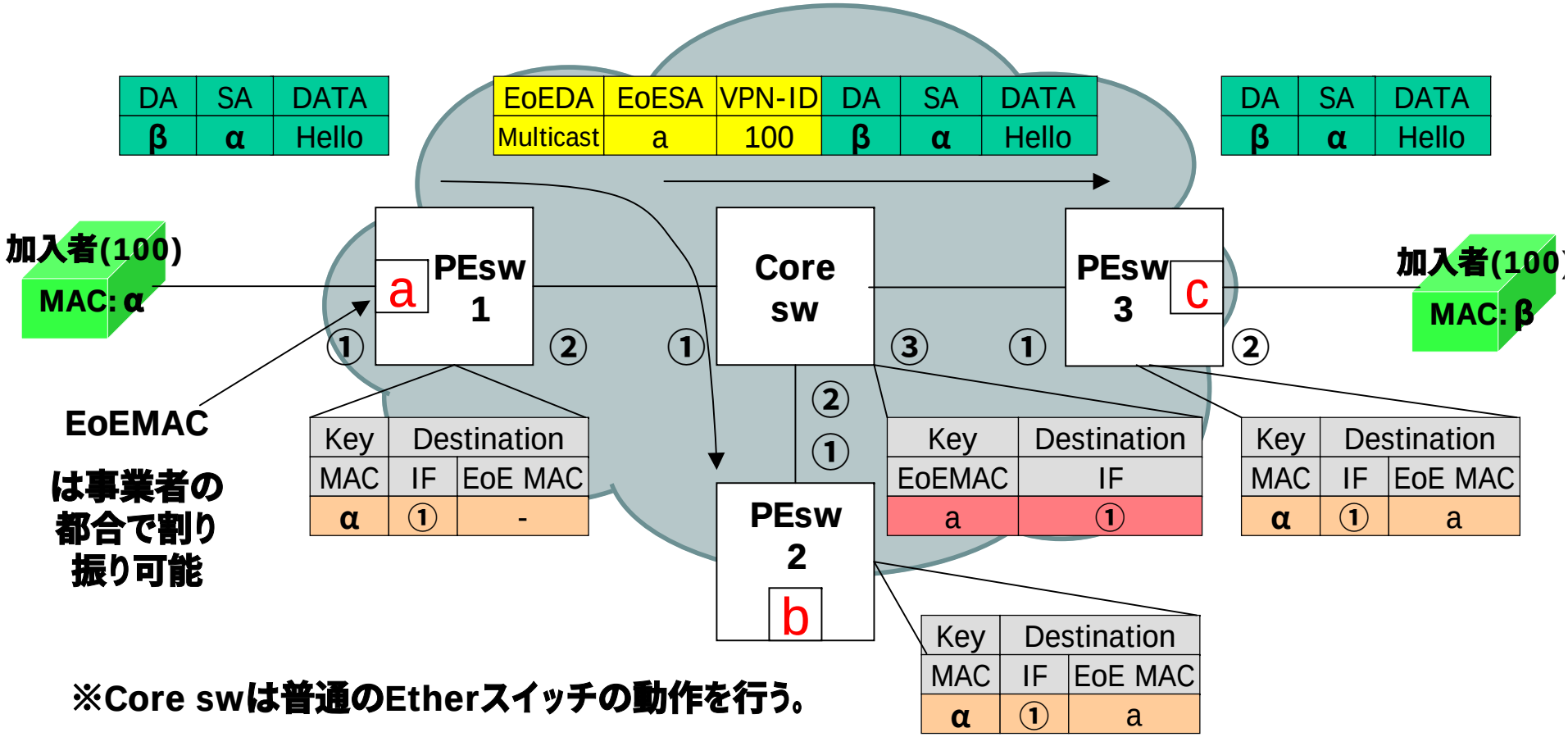


- ループ防止の観点から見たEoEの特徴
  - 階層化EoEMACアドレッシングとフィルタによるループフレーム破棄。
  - TTL導入によるループフレーム無限増殖防止。
  - TTLを用いたループ発生ノードの検出。

EoE(Ether over Ether)基本動作(1)

FDBにMACアドレスが学習されていない場合のEoE上のフレーム転送

- イングレス（入力側）PEにてMACアドレスが学習されていない場合、PEはEoEDAにマルチキャストアドレスをセットしフレームをフラッドする。

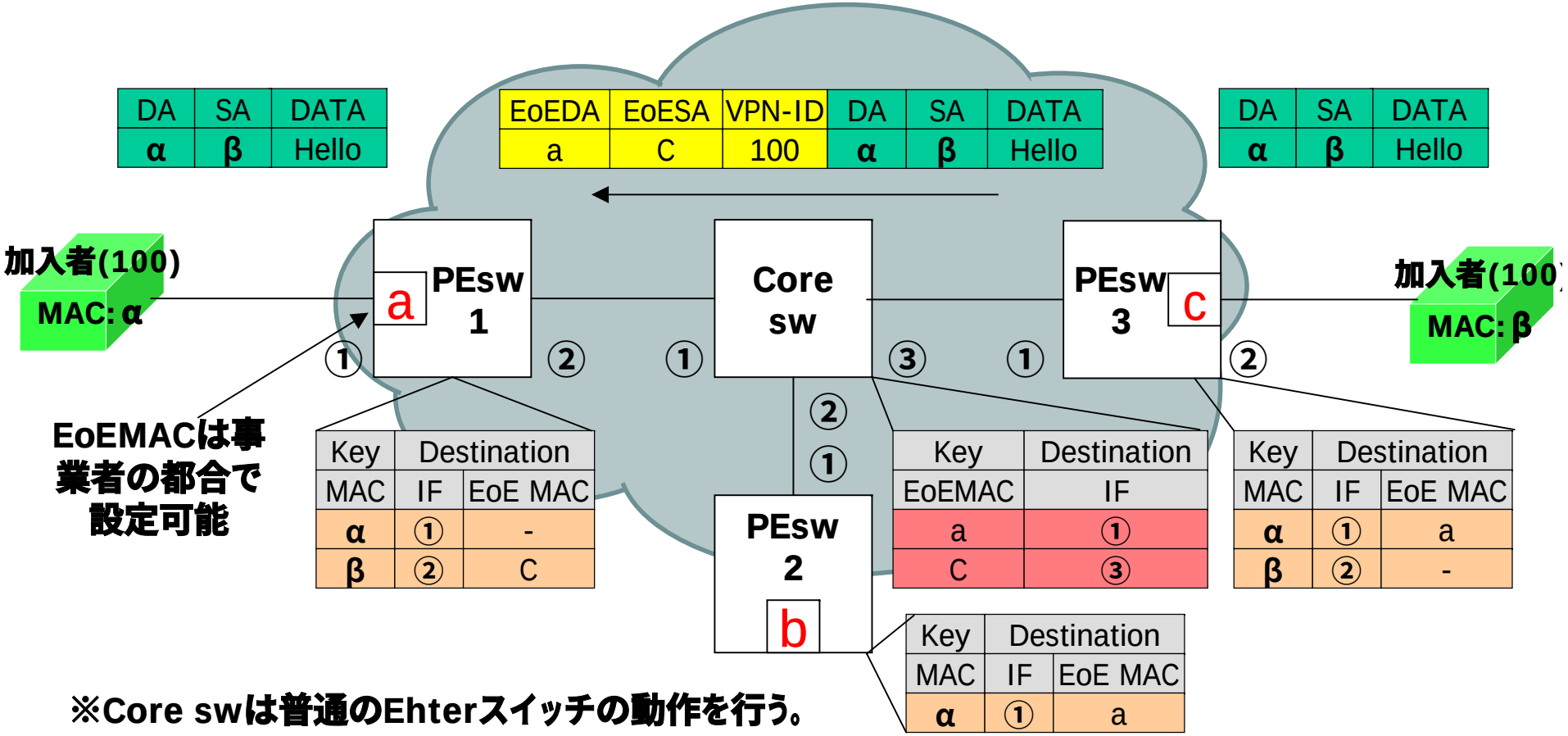


※Core swは普通のEtherスイッチの動作を行う。

EoE(Ether over Ether)基本動作(2)

FDBにMACアドレスが学習されている場合のEoE上のフレーム転送

- FDBにMACアドレスが学習されている場合は、学習の内容にそって、フレームが転送される。

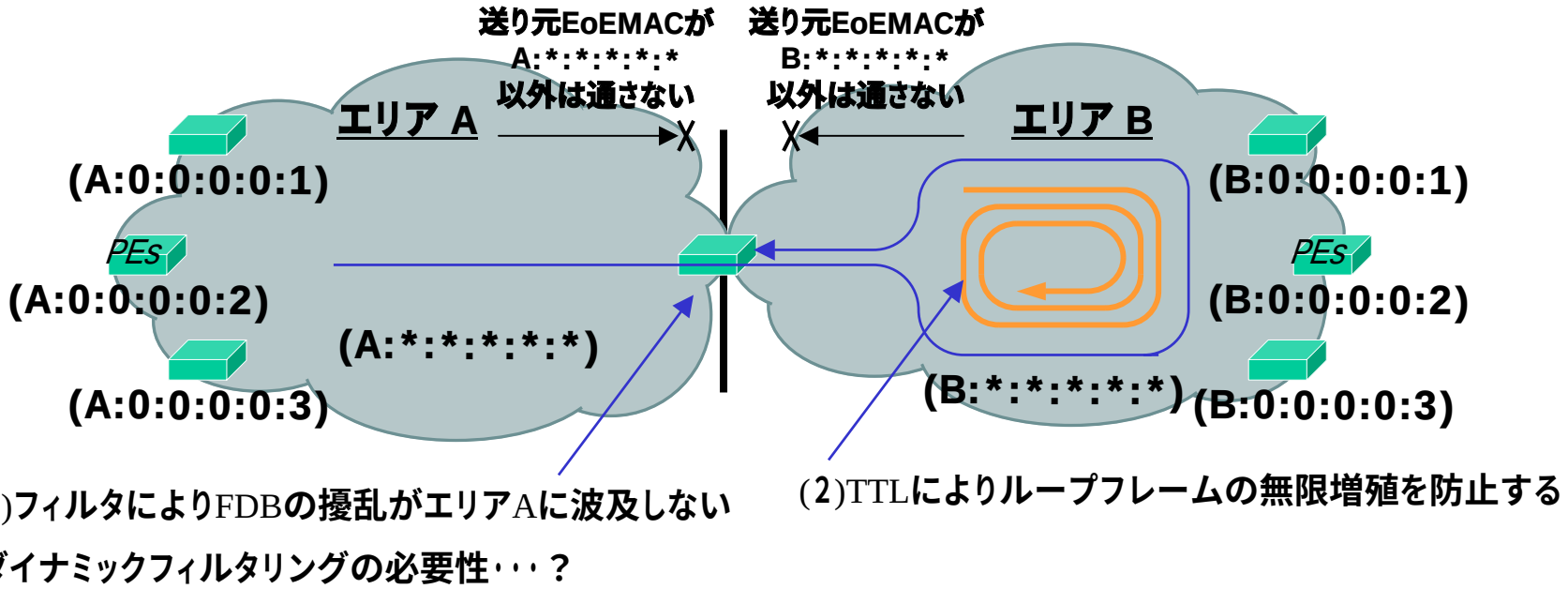


※Core swは普通のEhterスイッチの動作を行う。

EoEによるループフレーム破棄機構

（1） EoE階層化MACアドレッシングによるループ防止

エリアごとにEoEのMACアドレスを階層的に割り当てておき、中継ノードはフレームの送り元EoE MACアドレスを調べる事によりFDBの擾乱を発生させるようなループフレームを検出し、破棄する。



ダイナミックフィルタリングの必要性・・・？

（2）TTL導入によるループフレーム無限増殖防止

EoEのフレームにIPパケットと同様のTTLを導入しノードを経由するごとにカウントダウンさせTTLが0になるとフレームを破棄する機構を導入した事により、ループフレームの無限増殖を防止している。

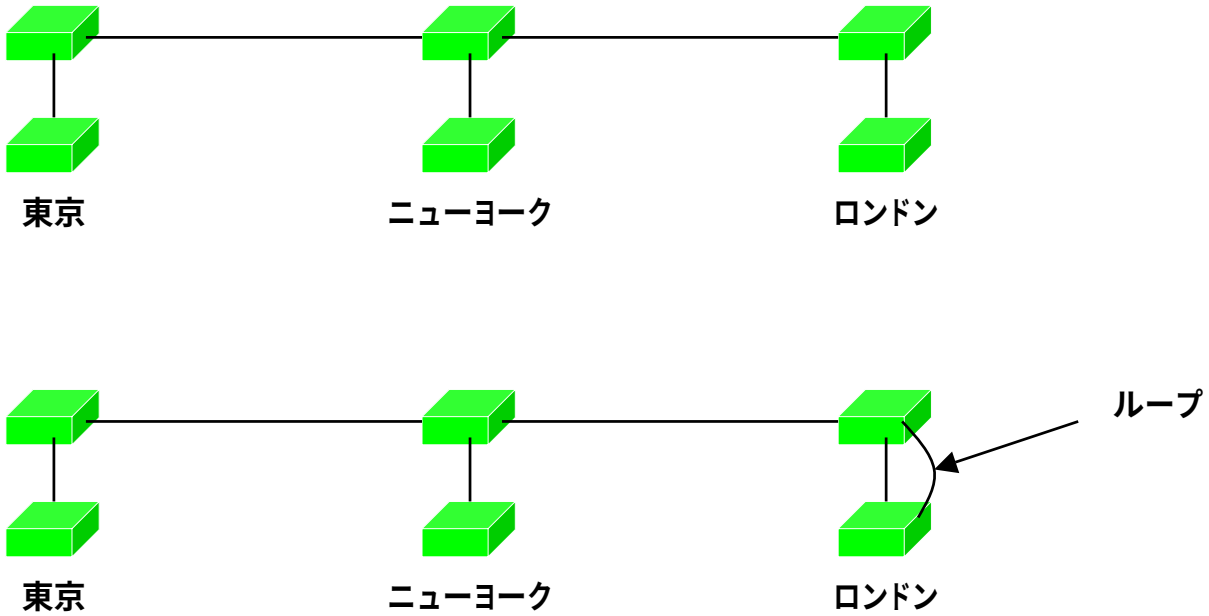
ループ位置検出の問題（１）

- EthernetなどのＭＡＣブリッジングにおいてループを検出する方式はいくつかある
  （１）ＦＤＢの書き換わり回数によってループを検出する方法。
  （２）マルチキャストなどのフレームを一定方向に送出しそれが戻ってくるのを観測する事によりループの方向と発生を検出する方法。
  （３）トラフィックに含まれるフラッディングトラフィックの量をモニタする方法。
  （４）ＴＴＬをexpireが発生する事によりループの検出を行う方法（拡張イーサネットにおいて）

これらの方法ではループが発生した事の検出は行えるが実際にどのノードがループしているかは判定出来ない！

ループ位置検出の問題（２）

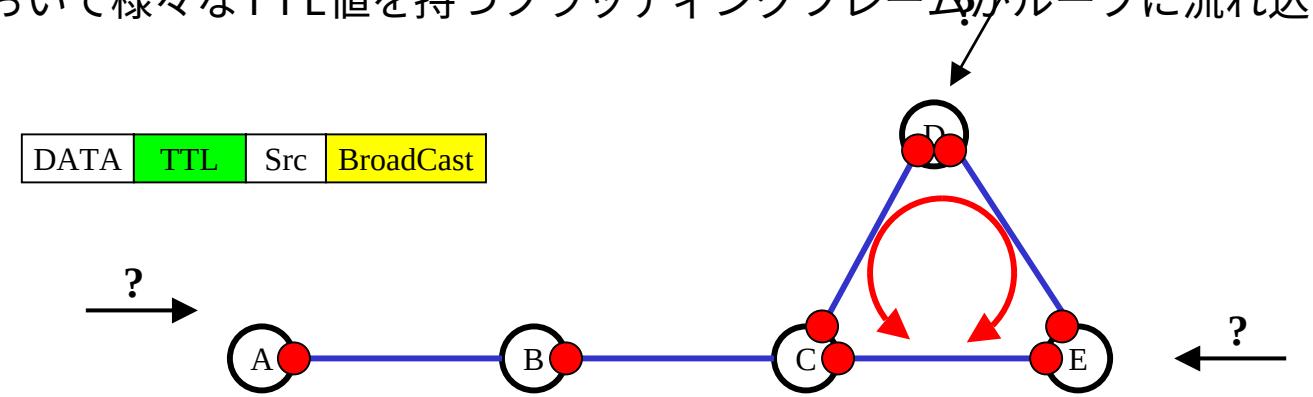
・ループの発生位置を検出するのは結構大変



ループが発生するとループを形成しているノードを発信元として、増殖されたマルチキャストやブロードキャストが多量に送信される為、ループ発生後に多量のトラヒックが流れ出す方向に探索を進める事によりループ箇所を検出する事が出来るが、スイッチの数が多いと時間がかかる。

EoEによるループ位置検出

普通のイーサネットではループが何処で発生しているのか検出するのは至難の業
ＥｏＥ網において様々なＴＴＬ値を持つフラッディングフレームがループに流れ込んでくると・・・



２ポート以上でＴＴＬ＝０による廃棄が発生するとループノード

- ループを構成するノードでは、２つ以上のポートから入力されたフレームにてＴＴＬ Expireによる、フレーム廃棄が発生する。（ループを構成するノードでは右回りのフレームが流入するポートでも左回りのフレームが流入するポートでもＴＴＬ Expireによるフレーム廃棄が発生するため）

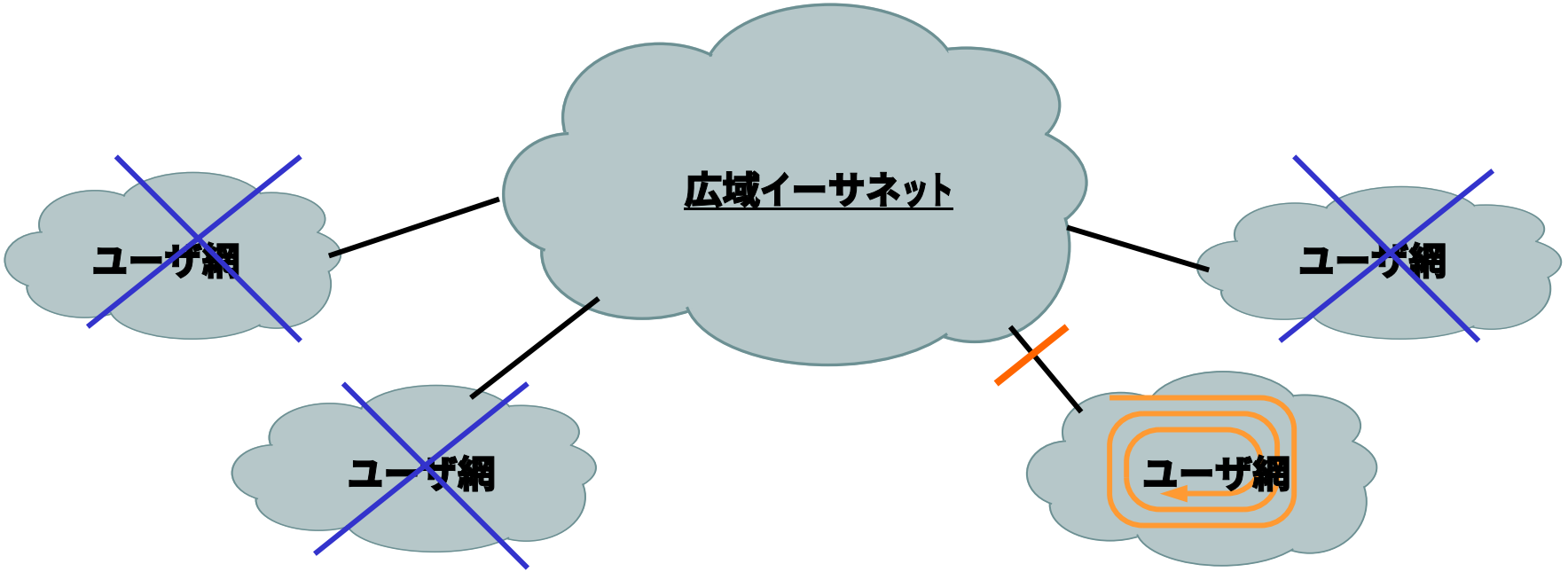
１ポートだけでＴＴＬ＝０による廃棄が発生するとループノードではない

- 直接ループを構成しないノードでは、ループの発生方向に向いたポートに入力されるフレームにてＴＴＬ Expireによるフレーム廃棄が発生する。（ループ方向から入力されるフレームのみが、ループ内を回るうちにＴＴＬを減らしている可能性がある為）

**通信事業者網内でのループであればこのトラップの上げ方を
変える事により瞬時にループしているノードを検出する。**

ユーザ＆アクセスループ対策

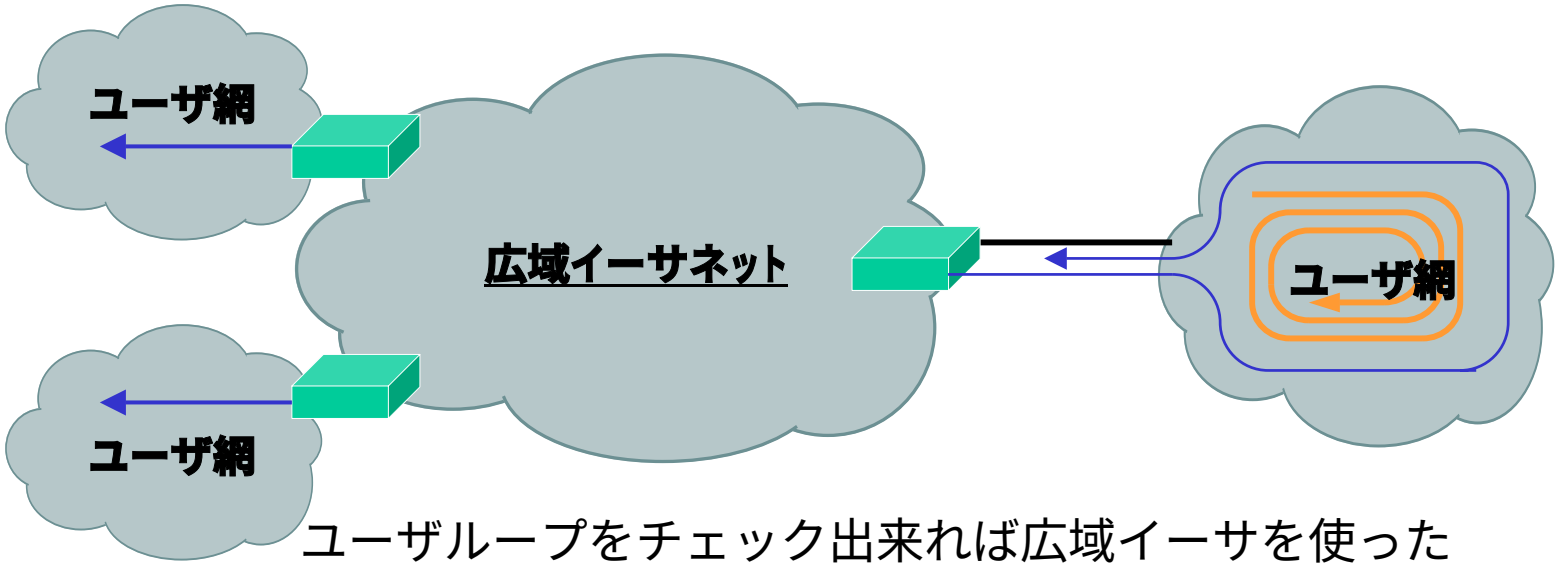
- ユーザループ対策の必要性
  - 広域イーサネット網内部でのループはほとんど防止出来るようになったが・・・「ユーザ宅内でのループ」や「専用線部分での折り返し」によってループが発生し疎通不能となる場合のケアは進んでいない。
  - イーサネットは、どの部分がループしても、全体が影響を受ける。
  - 全体が止まるぐらいなら、ループした拠点を切り離したほうがまし？



ユーザ網1つがループしたら同じセグメントの他のユーザ網も疎通不可となる。

ユーザ＆アクセスループ対策

- ユーザ網に対して定期的にループ検出用のフレームを流して、ループの発生を検出したい。場合によってはポートを自動閉塞（NTT研究所　鈴木宗良氏 loop-detection protocol -> IEEE…標準化は困難？）



ユーザループをチェック出来れば広域イーサを使った
システムのアベイラビリティ向上が図れる

相談です

ユーザと事業者とメーカのコンセンサスが必要！！

コンセンサスを作りましょう！！（標準化に関わってる人もよろしく）