

ホワイトボックス伝送の動向と 商用利用について

mixi, Inc.

Toshiya Mabuchi

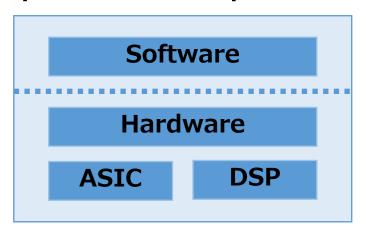




WhiteBox伝送装置?

ホワイトボックス伝送 **White Box Switch + Transponder**

Open Transponder



ASICに加えてDSPが加わったWBS

一般的によく語られるメリット

・ベンダロックインからの解放

Hardware / Software Disaggregatedによる選択肢の柔軟さ

・コントロールプレーン選択の自由

・ 自社のネットワークアーキテクチャに適合する形に組むことができる

・コスト最適化

・ エンジニアリングリソースと機器コストの最適化

本日の内容

• WhiteBox伝送を取り巻く動向と GoldStone(NOS)のアーキテクチャ NTT Electronics America 石田

Whitebox伝送とOSS NOSを 商用導入した背景/課題/展望 株式会社ミクシィ 馬淵



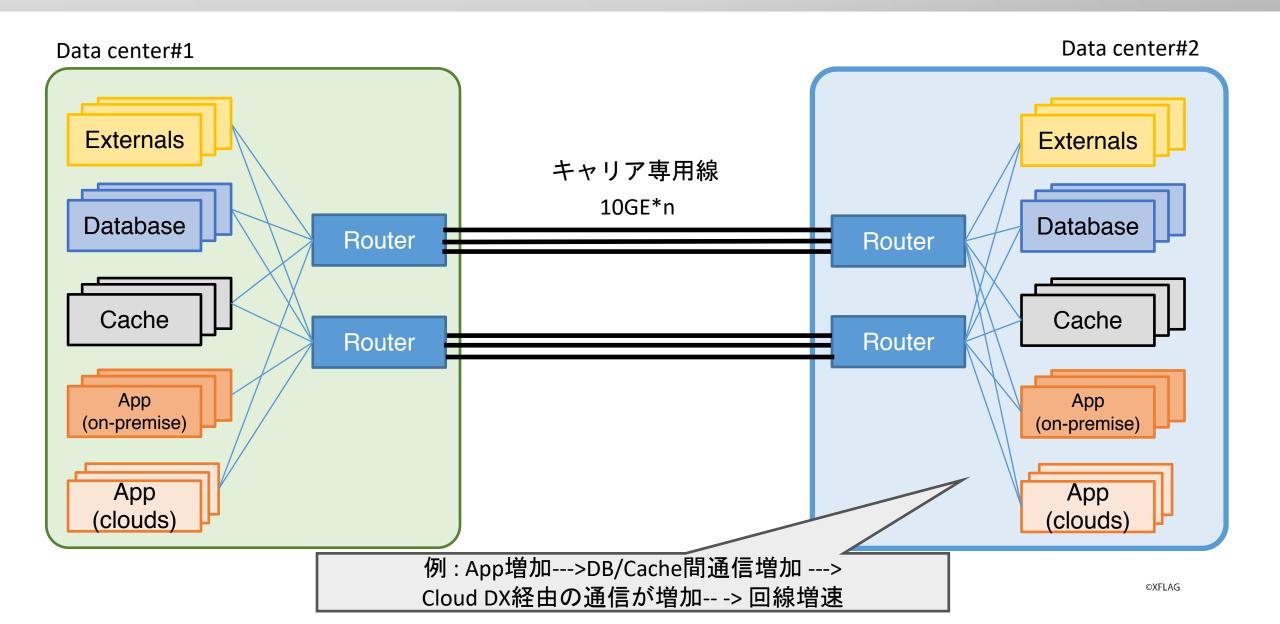
Cassini + Goldstone DCI use case and challenges

mixi, Inc.

Toshiya Mabuchi



mixiのDCI事情



DCIにおける課題

・回線調達のリードタイムの長さ

- ネットワーク拠点の追加や変更がそれなりの頻度で発生
- 納期は 2~4ヶ月
 - 拠点追加や増速の度にオーダーを走らせるのは辛い

・専用線のコスト増

- 10G * n回線でコストが線形増加
- 増加するコストの勾配の抑制をしたい

DWDM導入の検討へ

DWDMの運用上における懸念

懸念事項

DWDMオペレーション経験 :無し

管理対象機器 : 増加

ベンダーの既存機器 : 高価



達成したい要件

Server like Operation Cost Optimization Migrative for reduction : Kernel以降は自由に構成

: 安価&柔軟(ロックフリー)

:機器削減を見込める

Gold Stone

GoldStoneを採用したもう一つの狙い

・SONICの運用知見の蓄積

将来的なコアネットワーク利用のための助走 L3導入前に課題を運用しつつ出しておきたい

SONiC as Transponder

- ・派手にConfigを変更する必要はない
- ・最低限の機能利用からスタートして知見を蓄積できる
- ・問題が起きても上位レイヤーで救いやすい

ナレッジ蓄積のためにも、丁度いい箇所から初められる

検証フェーズ

Cassini + GoldStone検証

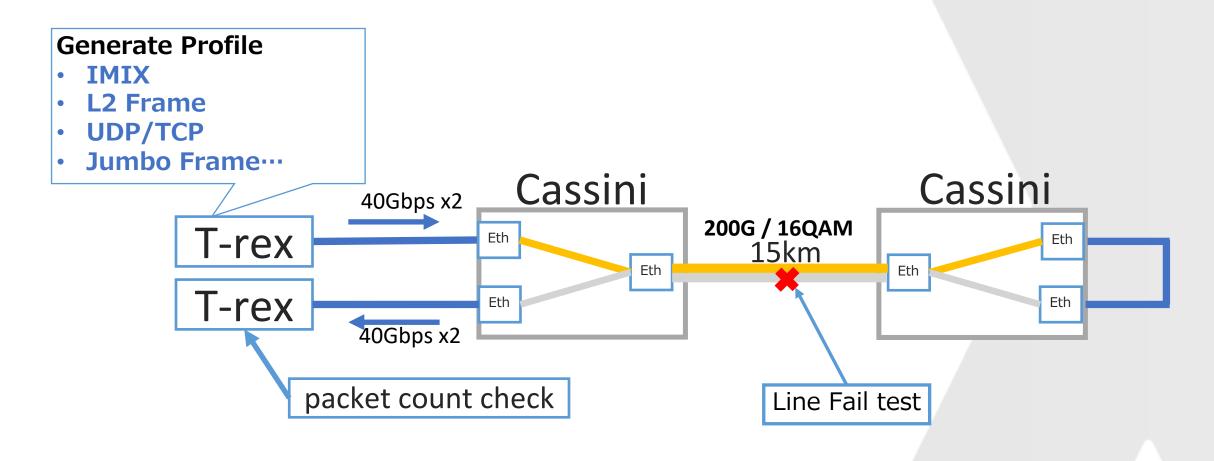
• 安定性

- サービスと同等の状況でリソースが安定し、ロスなく伝送するか
 - Monitoring
 - Memory leak/ CPU load / Temperature / DiskIO / DSP fec,rms
 - Fowarding test
 - 商用を想定したパケットを流し続ける
 - Loss / Error check

• 運用フロー

- デプロイAnsible整備
- SONiCとDSP Metrics監視用Prometheus Exporter実装&整備

テスト時の構成図



Over 6month

検証過程での課題

課題1: Goldstone ≠ Transponder

・リンクダウン転送がない

- CassiniはDSPを搭載できるASIC搭載のスイッチ
 - LINE側にもASICを持つ
 - Clientポートが落ちてもLINEは落ちない
 - つまり、ポートのdown転送は行われない



CoreのIGPのdown検知がHold-timer依存になる

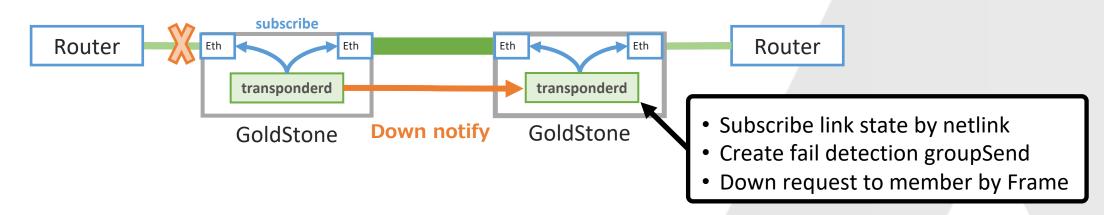
P2P DWDMとして使うには、そのままでは適さない

解決策: BFDによる断時間の削減

現状: 両端のルータでBFD 100msec*3



今後:より高速&低負荷に検知するための "transponderd"を開発&デプロイ



課題2: SONiC synd is delicate

・SONiCの手動のオペレーションは相当厳しい

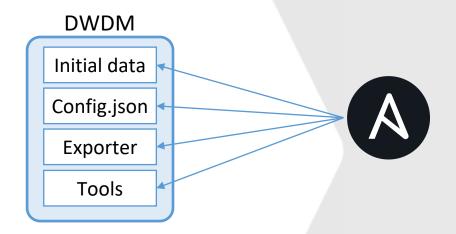
- Not Day2 config operation
- Typoが混ざるとSyncd(ASIC controller)等down
 - Configに対してのValidationが甘い

Typo in configuration

```
2019-09-12.06:37:27.491030|s|SAI_OBJECT_TYPE_PORT:oid:0x100000000013|SAI_PORT_ATTR_ADMIN_STATL=false
2019-09-12.06:37:27.491149|s|SAI_OBJECT_TYPE_PORT:oid:0x100000000013|SAI_PORT_ATTR_SPEED=400000
2019-09-12.06:37:27.491252|s|SAI_OBJECT_TYPE_PORT:oid:0x100000000013|SAI_PORT_ATTR_ADMIN_STATE=true
2019-09-12.06:37:27.493167|n|switch_shutdown_request||
```

解決策: Day1 Configで完結

- ・Day1 Config(初期投入)で全て完結させる
 - 伝送用途なら必要十分
 - AnsibleでConfig / Tool群を投入



- ・コアでL3を任せる場合には、Validationは流石にほしい
 - Configの<u>Validationを行うPR</u>がMerged!!
 今後は改善されるかも

課題3:様々な起因でOSクラッシュ

・温度管理ミスによるCrash

- CassiniはBMCが非搭載
- OS上でFANコントロールが必要

SNMP等,マネジメント関連のリソースのメモリリーク

- SONiCは枯れていないOSS
 - 仕様変更に伴ったバグも多く出る
- マネジメント系であっても 相応の検証は必須



解決策: リソース監視を徹底

・リークや不具合は起こり得る前提

- サービスの担保は全体で行う
- 原因をはっきりさせるためにMonitoring項目を強化
 - Metricsをコード上で抽出する手段が用意は多くあるので、 そこまでキツくはない

Prometheus expoterでまとめて取得

・基本リソース : <u>node_exporter</u>

• TAI特有Metrics : gRPC経由で取得

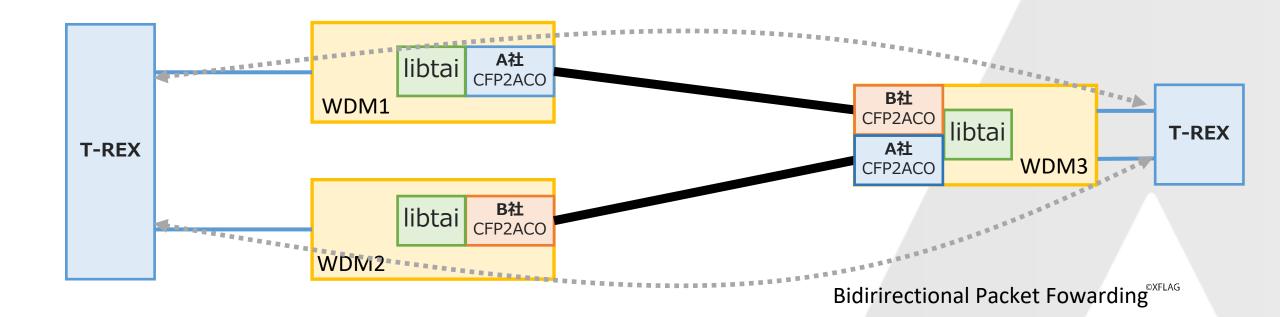
• **SONiC固有Metrics** : Redis経由で取得

各センサ温度管理 : ONLP lib 経由で取得

Ex. 相互接続性検証

・CFP2-ACOの相互接続性試験

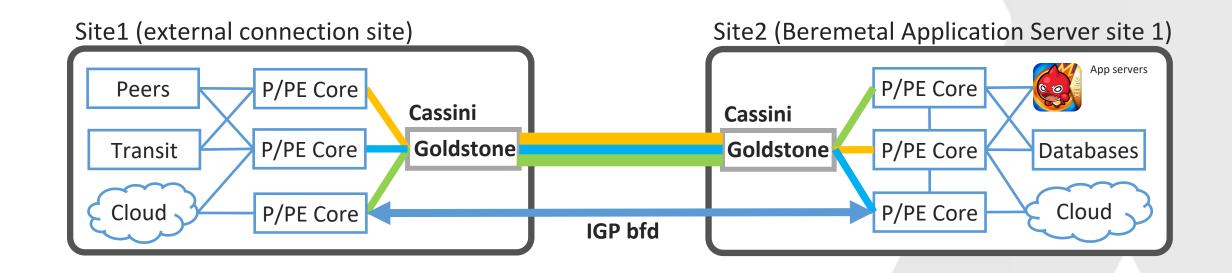
- A社 / B社の2メーカーで試験
- ・光レベル,FEC,確立後のパケット伝送、全て問題無し
- TAIがマルチベンダーもカバー



商用導入

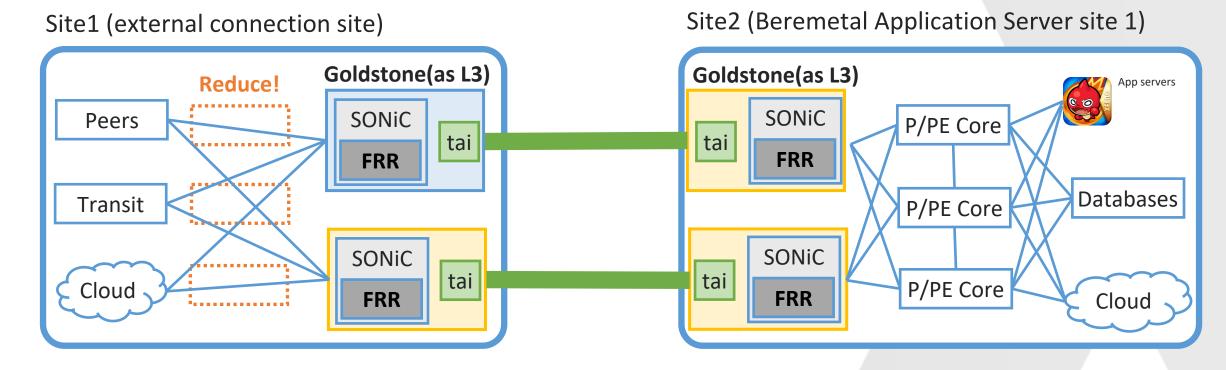
Current Cassini + Goldstone use case

- PointToPoint DWDM
 - BFDによって断時間を保証



WIP: Migrate include Layer3 Routing

- 特定拠点のルータ機能吸収
 - GoldStoneだけで完結させれる点はルータとして扱う
 - 管理対象を削減し、運用の最適化を実施



WhiteBox DWDM + OSS NOSのメリット

・コスト

特にCFP2ACO等の単価

機器のポテンシャル

伝送装置の役割だけで終わららない その後のL2/L3 Migrationも可能にしていける構造

・共通技術の活用

WDM / Server / Switch 全てで共通の技術を活用

Core Component

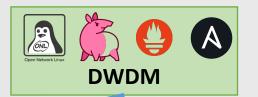
- ONL, SONiC, FRR, Zebra, SAI, TAI etc...

Automation/Monitoring

- Prometheus, Ansible, gNMI, Go/Python tools...

→ Network特有の運用を削減

















WhiteBox DWDM + OSS NOSのデメリット

・導入のナレッジ蓄積に時間がかかる

- SONiCを初めから入れてる企業ならOK
- DWDMのためだけにSONiCを運用まで持っていくのは辛い

・エンジニアリングリソース

• 特にトラブル解析時に多く必要となる

・「手作業でオペレーション」は厳しい

• 初めからある程度の作業不要にする取り組みは必要

Summary

- ・ Cassini + GoldStoneをプロダクションに導入
 - Point-to-PointのDWDM
 - SONiCをDWDMから運用開始して他機器に展開するアプローチ
 - コスト面では専用線より優位
 - サーバーライクなオペレーション
- ・SONiCをより上位レイヤーで活用するための試験を実施中
 - SONiCのL3 Routingを本格的に利用して機能統合

Copyright © 2019 mixi, Inc.

©XFLAG

議論したいこと

- ・ そもそも、ダークファイバ運用していますか?
 - してない方は、どのようにDCIを実現していますか?
- ・ 伝送のオープン化について
 - ホワイトボックス伝送装置に興味ありますか?
 - 安価に便利に伝送する他手法との比較 (10G-DWDM, 100G-PAM4-DWDM, 400G-ZR, CFP2DCO on Switch)
 - ルータから伝送を吸収するアプローチ
- ・ 光伝送の運用で課題だと感じている点
 - コスト
 - 取り回し
 - 運用的知見の不足
- ・ SONiCのようなOSSのNOSを商用網で利用することについて
 - 成長途中なOSSとの付き合い方
 - 商用サポートvs取り回し,拡張性
 - 使いたいけど、実装まで手は出せない?

Thanks

