

# システム冗長性向上を目指したネットワークインフラの導入 について

梶田 秀夫<sup>1,a)</sup> 村田 和義<sup>1</sup> 渋谷 雄<sup>1</sup> 黒江 康明<sup>2</sup>

**概要:** 京都工芸繊維大学では 2012 年 12 月に, 10 年以上使用してきた 1000BASE-X を基盤としたネットワークを更新し, 10GBASE-X を基本とするネットワークインフラ KITnet5 を導入した. KITnet5 は, イーサネットとファブリックの両方に対応した Brocade Networks 社のボックス型スイッチである VDX 6720 を集約スイッチとして使用した. また, 約 10km 離れた松ヶ崎キャンパスと嵯峨キャンパスを, SINET データセンター経由で 10GBASE-ZR を用いて接続した. これにより, 複数キャンパスを跨る SAN を構築することができるようになり, コンピュータシステム自体を複数キャンパスに跨って構築が行える基盤となった. 本報告では, 本構成を導入するにあたっての経緯や, 利点・欠点および一年間の運用経験などについて述べる.

**キーワード:** 情報インフラストラクチャ, 10Gbps, イーサネットファブリック, 拠点間通信

## Experience of High-Speed Network Infrastructure for Computer System Redundancy

HIDEO MASUDA<sup>1,a)</sup> KAZUYOSHI MURATA<sup>1</sup> YU SHIBUYA<sup>1</sup> YASUAKI KUROE<sup>2</sup>

**Abstract:** In our university, we have updated our network infrastructure from Cisco based 1-gigabit Ethernet system to Brocade based 10-Gigabit Ethernet system in December 2012. We use Brocade VDX Switch for aggregation Switch. It enables not only IP networking but also providing Fibre Channel and/or Fibre Channel over IP/Ethernet. Moreover, it provides a connectivity between Matsugasaki and Saga campus (far away about 10km) with 10GBASE-ZR Ethernet via Data Center of our region. It enables a Campus-wide IP Storage Area Network (SAN), then computer system can be operated with distributed environments, so our computer system will have redundant feature against disasters. In this paper, we show merits, demerits and experience of our system.

**Keywords:** Information Infrastructure, 10Gbps, Ethernet Fabric, Inter-Campus Network

### 1. はじめに

京都工芸繊維大学 情報科学センターでは 2012 年 12 月に, 10 年以上使用してきた 1000BASE-X を基本としたネットワークを更新し, 10GBASE-X を基本とするネットワー

クインフラ KITnet5 を導入した.

KITnet5 は, Brocade Networks 社の L3 スイッチである CER2024 を基幹ルータ, Energy Efficient Ethernet (EEE) 標準に対応した L2 スイッチである ICX6450 をエッジスイッチとするオーソドックスなスター型構成の中に, Virtual Cluster Switching (VCS) Fabric 技術が搭載された VDX6720 を集約スイッチとして展開した構築をしていることが特徴である. また, ファイアウォールとして, Fortinet 社の Fortigate 1000C をキャンパスの 2 拠点に分散配置して広域な冗長構成をとっていること, ALAXALA 社の

<sup>1</sup> 京都工芸繊維大学 情報科学センター  
Center for Information Science, Kyoto Institute of Technology

<sup>2</sup> 京都工芸繊維大学 情報工学部門  
Information Science, Kyoto Institute of Technology

a) h-masuda@cis.kit.ac.jp

AX3650S(NEC IP8800/S3650)を用いた Shibboleth 対応の captive portal を持つこと, Cisco 社の AIR-CT5508 と学内に配置した 70 台以上の AIR-AP1131AG, AIR-AP1142N を用いた eduroam にも対応した無線 LAN インフラを持つことも特徴である。

## 2. システム更新の目標

### 2.1 本学におけるネットワーク整備の現状

ネットワークインフラは, 教育, 研究から事務作業に至るまで, いまやなくてはならないものとなってきている。しかしながら, 大学においては, 定期的に更新が可能なレンタルもしくはリース型のシステムとして導入されているところは少数派であり, 買い取り型の調達であることが多い。本学でも, 1994 年に FDDI (Fiber-Distributed Data Interface) を基盤とするネットワーク (KITnet1) を整備し, 1996 年には ATM (Asynchronous Transfer Mode) ネットワーク (KITnet2), 2001 年に整備されたギガビットイーサネットワーク (KITnet3) を, いずれも補正予算による買い取り型で整備している。

2009 年には, KITnet4[1], [2] として, KITnet2 の部分を置き換える整備をし, KITnet3 と KITnet4 の併用運用を開始した。しかしながら, KITnet3 を構成する機器の保守期限が切れ, 早急な更新が必要となっていた。

### 2.2 本学における更新の計画

情報科学センターとしては, 本学のネットワークシステムは重要なインフラであり, 不定期の買い取り型の調達を今後も続けていくことは重大なリスクであると主張し, 本学の設備マスタープランでは, 8~10 年での更新を計画するように計画に組み入れられている。また, レンタル型であっても買い取り型であっても, いつかは物理的な機器の入れ替え作業が発生し, ネットワークの一時的な停止は避けられない。しかし更新にあたって, ネットワークシステムのすべての機器を一気に入れ替えるような構成となっていては, 長期に渡る停止となってしまう, インフラとしては大きな問題 (リスク) を抱えてしまうことになる。

従って, KITnet4 の更新から, ネットワークインフラは, 半分ずつ更新できる構成にし, 併用期間中に徐々に移行できる構成 (チクタクモデル) にすることとしている。

## 3. 前回の更新時の反省点

### 3.1 リング状トポロジのイーサネットの弱点

KITnet4 では, リング状のトポロジを採用した。リング状トポロジの利点は, 高価な GBIC を大量に持つコアスイッチ機器が不要となり, かつ, システム全体の GBIC 数も削減できることであり, KITnet4 ではそれが非常に有効に作用した。

一方で, リング上の機器の単体故障ではなく, 複数の機

器もしくはリンクの故障が発生すると, ネットワークの分断が発生する。KITnet4 では, VC (Virtual Chassis) と呼ばれる仮想化技術によるリング構成を採っていたため, 分断による VC 機能の再構成機能が想定通りに働かない場合が散見された。

### 3.2 仮想化技術による長時間の一時停止

ネットワークインフラ部分は比較的低いレイヤー, とりわけレイヤー 1 は物理的なデバイスが絡む部分であり, 物理デバイスに関する障害検知機能は重要な役割を果たす。一方仮想化は, いわば物理デバイスの差異を吸収しそれを見せなくする技術であるため, 物理デバイスの情報は隠蔽されがちであり, 障害判別が難しくなってしまう部分が存在する。さらに, Juniper 社の JUNOS は, BSD をベースとする安定性を重視し, 多くのノウハウが蓄積されたネットワークデバイス用の OS と言えるが, VC 自体がまだまだ新しい技術であるため, ノウハウの蓄積が不十分であるところも否めない。

例えば, VC を構成する機器との接続に使用されるポートでの障害が, VC 自体の安定性に強く依存してしまうため, 単なるポート不良よりも障害判別が難しくなる場合があったり, VC を構築した機器数に応じて設定の反映 (commit) にかかる時間が無視できない程度 (2 分程度) になることがある。また, JUNOS は, ufs ファイルシステムを内部に持つ UNIX システムであるため, UNIX の知識を有するオペレータには非常に親和性が高いが, ブートシーケンスやシャットダウンシーケンスを守る必要があるため, 不意の電源断に弱い部分がある。その為, 本学の環境では, キャンパス内の建屋スイッチの設置場所すべてに UPS などを入れることは難しい上に UPS と連携して停止する機能なども無いため, 時々起動に失敗する場合がある。実際, dual root filesystem を採用して軽減策をとっているが, それでも停電明けには何かしらの不具合が発生することが無くなってはいない。

それでもなお, 多数のスイッチデバイスを個別に管理するのではなく, 一体として管理できる機能は非常に有用である。

## 4. 更新計画

2008 年度予算で大容量基幹情報ネットワークシステム一式の調達を開始した。この調達は買い取り型であった。

### 4.1 技術的要件

調達にあたっては, KITnet4 と同様に, 以下の観点での選定を行った。

- (1) 導入時点でできる限り最先端技術を採用する。
- (2) できるだけオープンな技術を採用する。
- (3) 最低限の冗長化技術を採用する。

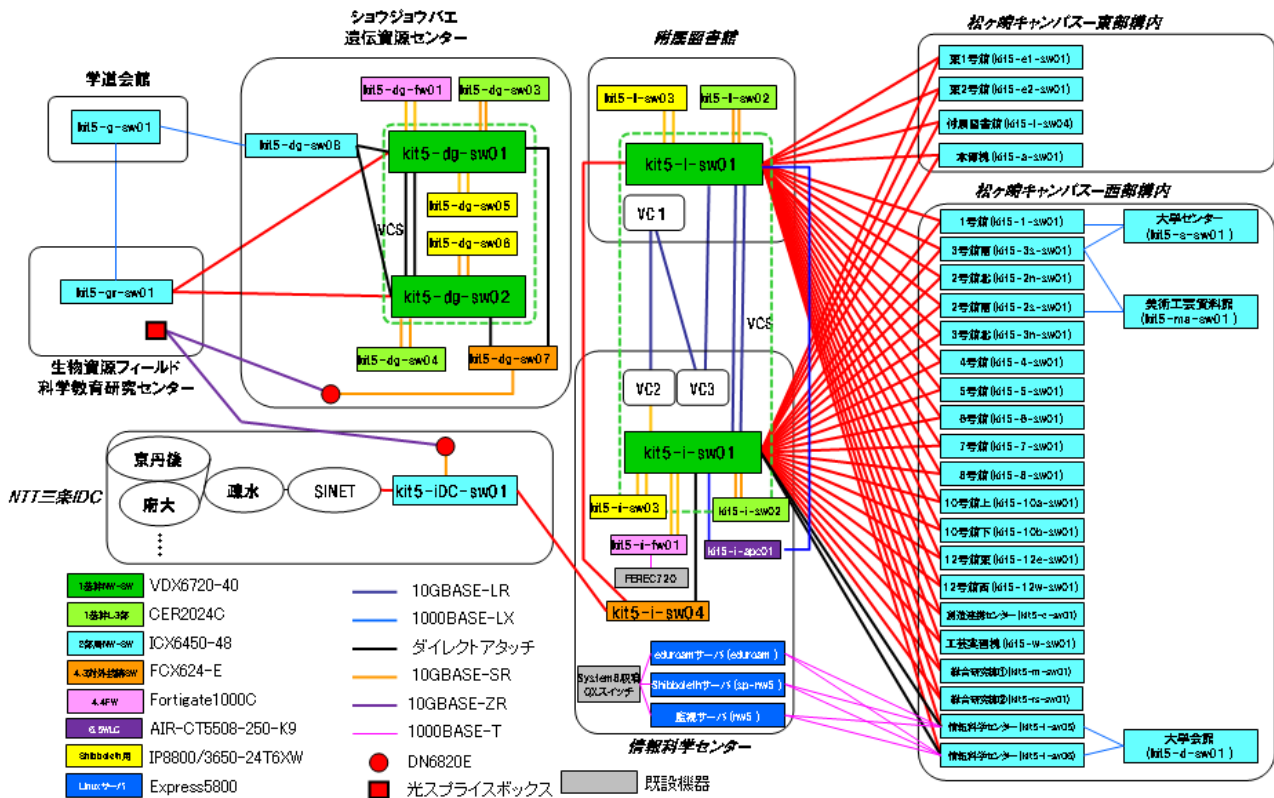


図 1 システム構成概略図

- (4) 性格の異なるシステム構成を組み合わせた構成を採用する。
- (5) できるだけ同一の機器で揃える。  
これに加えて、以下の要件を含めた。
- (1) 松ヶ崎と嵯峨を大容量に接続する。  
東日本大震災を経験し、システムの分散配置に対する関心が高まっており、重要なデータの保全に対する要求が上層部からも出てきていた。一方で、データをクラウドサービスのような第三者組織に安易に預けることは、昨今の情報漏えい事件が多く発生しており躊躇われるところである。そこで、本学の松ヶ崎キャンパスと嵯峨キャンパスを大容量回線で接続することで、データ流通基盤を整備することを要件とした。
- (2) 松ヶ崎と嵯峨で相互補完できるようにする。  
いずれかのキャンパスだけでもネットワークサービスが継続できるようにする。そこで、できるだけ冗長構成をとるが、機器の設置場所を分散設置できることを要件とした。

### 5. システムの構成

本システム (KITnet5) では以下の技術を用いてシステムを構成した。

- (1) 10GbE (10GBASE-LR) を基盤とするバックボーンとする。
- (2) 放射型トポロジを基本とする 3 層構造 (core-aggregation-edge) とする。
- (3) Brocade VCS Fabric 技術を用いた広域ストレージ基盤採用する。
- (4) KITnet4 とは OSPF を用いて相互接続を実施する。

図 1 は、本システムの構成概略図である。  
KITnet5 は、Brocade Networks 社の VDX6720 をアグリゲーションスイッチとして利用し、コア Layer3 スイッチとして同社の CER2024 を\*1、各建屋のエッジスイッチとして ICX6450-48 を放射状に接続する構成となっている。

\*1 VDX6720 は、Layer3 スイッチ機能を有しているが、導入時点では IPv4 のみ対応であったため、CER2024 で IPv4/IPv6 両方の Layer3 スイッチ機能を提供する構成としている。

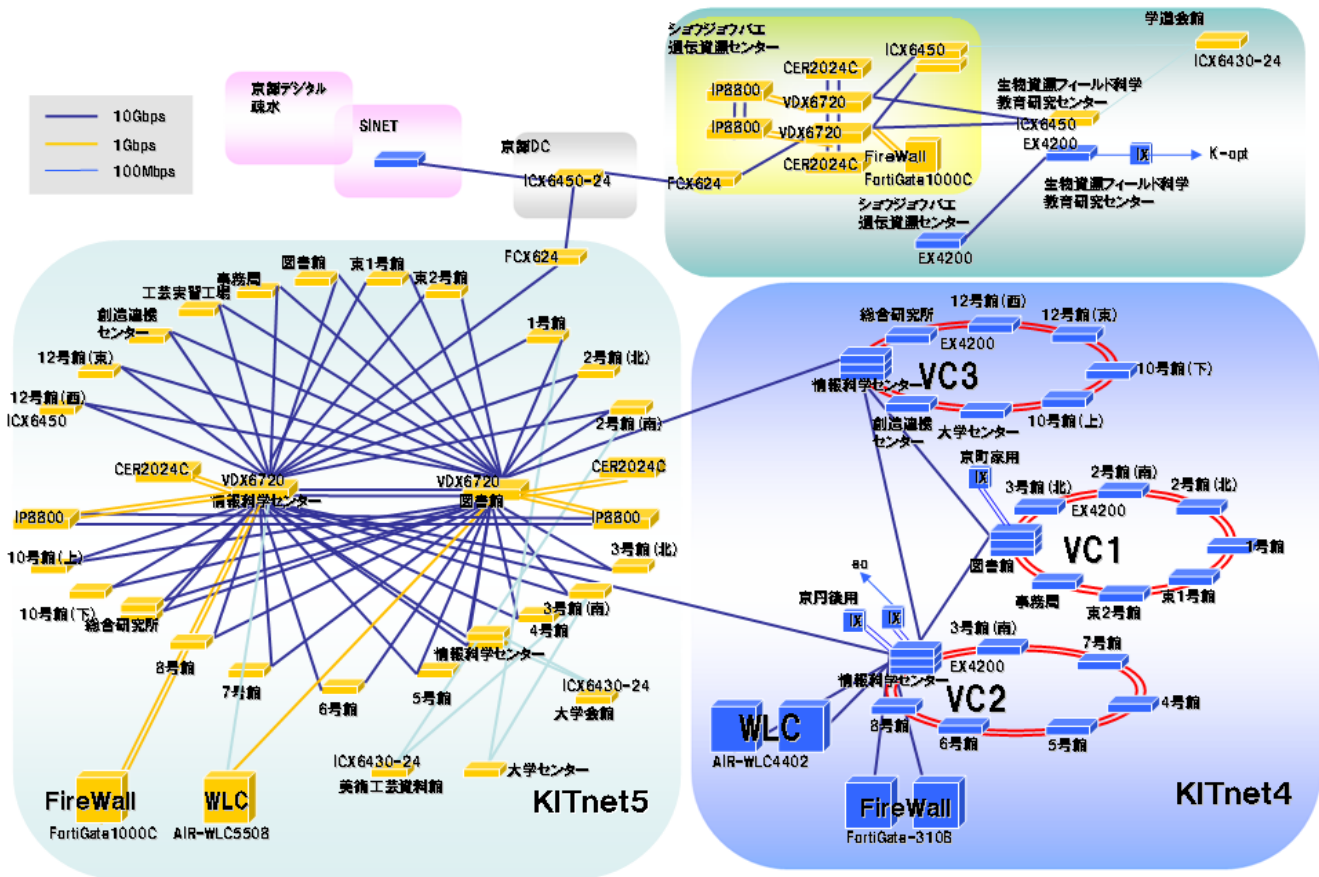


図 2 システム配置図

### 5.1 基幹装置の建屋分散配置

通常、キャンパスネットワークのコアスイッチは、センター建屋内で二重化するなどして冗長性を持たせる方式が多いと考えられるが、KITnet5では、コアスイッチおよびアグリゲーションスイッチを、松ヶ崎キャンパス内の約200mほど離れた情報科学センターと附属図書館に分散配置している。それぞれに分散配置したVDXはVCSを組んだ上でロジカルシャーシを構成しており、仮想的には一つのスイッチに見えている。さらに、各建屋のエッジスイッチとの接続は、単純なリンクアグリゲーションに見えるため、STPなどを使ったアクティブ・スタンバイとは異なり、両方の通信リンクをフルに活用できる。情報科学センターと附属図書館は、本学の特別高圧受電変電設備からみると、別システムのキュービクルの配下に接続されているため、学内の建物の工事に伴う停電で同時に止まることはないと考えられ、停止機会も軽減できる。

また、小規模な建屋にまで全て10Gbpsのリンクを準備することはコストパフォーマンスが低下する為、建屋を一次拠点と二次拠点に分け、一次拠点はVDXへのリンクアグリゲーションとするが、二次拠点はエッジスイッチへの

1Gbpsリンクにすることとした。ただし、冗長性確保の為に、二次拠点のエッジスイッチは、二箇所の異なる建屋にある一次拠点スイッチに収容することとしている。この接続には複数のICXではロジカルシャーシが組めないためリンクアグリゲーションは使えないが、Brocade社独自技術のプロテクトリンク機能を使用し、STPは使わないようにした。これにより、KITnet5としてはSTPフリーとなり、他のネットワークとの接続にSTPを使っても良いことになる。

さらに、学外とのルーティングおよびファイアウォール機器は、約10kmほど離れた松ヶ崎キャンパスと嵯峨キャンパスに分散配置した。ルーティングは、CER2024上のVRF (Virtual Routing and Forwarding) で構成したルータで行い、ファイアウォールは、Fortinet社のFortigate-1000Cをアクティブ・スタンバイ構成で配置している。これにより、受電設備点検で一方のキャンパスが停電したとしても、もう一方のキャンパスで学内外のインターネット接続を継続できるようになっている。

## 5.2 シリアルコンソールの相互接続

KITnet4 と KITnet5 は相互補完的な構成としており、主要な建屋には両方のエッジスイッチが設置されている (図 2)。更新前は、エッジスイッチの交換時に、一時的に他方のスイッチの下にカスケード接続する、といったことを想定していた。しかし、停電などからの復旧時に、KITnet4 の VC の再構築に失敗する事例に対して、コンソールアクセスが必要になる場面が何度か発生したため、なんとかならないか、という要望が運用現場から出ていた。

そこで、それぞれの機器のシリアルコンソールに、RaspberryPi をベースにした OBJECT 社の SCONVT-II を接続し、相手のネットワーク機器が生きていればシリアルコンソールに遠隔からたどり着けるようにした。

## 6. 現在の運用状況と考察

稼働開始後の所感は以下の通りである。

### 6.1 良い点

- コア系装置のコンパクト化。  
KITnet3 は Catalyst6506 を 2 台構成であったことに対して、CER2024 (1U)、VDX6720 (2U) を 2 セット有するだけで済んでいる。対外接続用のメディアコンバータなどは情報科学センターと附属図書館で別々に配置するわけにはいかないが、機器数が最小になったため、情報科学センター側に用意した 3000VA の UPS のみで最低限の機器の長時間運用が可能となっている。
- 建屋機器のハードウェア保守費の削減。  
一次拠点の ICX6450-48 および二次拠点の ICX6430 は、Limited Lifetime Warranty であるため、特に保守経費を計上する必要がない。

### 6.2 懸念、要改善点

- VCS Fabric は拠点間までは延長できない。  
VCS では、レイテンシに対する要求が厳しく、導入当初は 10km を超える接続には適用できなかった。現在は 30km まで延長できるように改善されているが、途中の機器がすべて VDX である必要があり、本学では適用できていない。本年度に更新した情報基盤システムでは、10Gbps のイーサネットリンクとして活用し、iSCSI や NFS による SAN 構築を実施している。
- VDX のレート制御の影響範囲が広い。  
VDX6720 は、ASIC 毎に入力キューと出力キューを持っている。内部でスイッチング後にどこかの出力キューで溢れが発生すると、その出力キューへの送分を捨てるのではなく、送るべき入力を持っている ASIC に対して、それ以上の入力が無いように入力を破棄する、といった動作をする。従って、どこかの出力キューが詰まってしまうと、多くの入力パケットが

捨てられてしまうことになる。導入時は、VDX6720 に 10Gbps リンクと 1Gbps リンクが混在していた為、大量のトラヒックが発生した際に、1Gbps リンクへの出力が溢れてしまい、それに伴って、いくつかの 10Gbps リンクの通信が止まってしまう症状が発生した。現在は、ファームウェアの修正に加えて、1Gbps リンクを VDX から無くすことで概ね解消しているが、できるだけ通信を止めたくない機器の接続ポートは、ASIC の物理的なポート割当を考慮して決定する必要があると考えられる。

- エッジスイッチの集中管理機能の欠如。  
KITnet4 の VC に比べて、すべてのエッジスイッチを個別に管理するため手間がかかる。しかし、これは、KITnet3 時代の個別管理と同等なので、改悪とは言いがたいが、改善が望まれる。
- ICX の loop detection 機能の能力不足。  
ICX のループ検知システムには、Strict mode と Loose mode があり、現時点では Strict mode で使用している。この場合、ポートごとに専用のフレームを送出し、それが戻ってくることでループを検知するので、複数のポートを短絡した場合は機能しない。Loose mode であれば、VLAN 単位で異常を検知できるが、検知時に物理ポートを err-disabled にしてしまうため、エッジスイッチ自体が切り離されてしまう。これについては、ファームウェアの更新により対応できるようになる見込みである。

## 7. おわりに

VCS Fabric 技術を利用可能とした 10GbE を基盤とするネットワークインフラとして、Brocade Networks 社の VDX6720 を採用して構築を行った。また、松ヶ崎キャンパスと嵯峨キャンパス間も 10GBASE-ZR を用いて 10Gbps の容量を持つネットワークを構成した。VCS Fabric 技術は、これからもネットワークインフラやサーバ環境、ストレージに至るまで適用される技術であり、今後も活用されていくことが予想される。現在、来年 3 月に更新予定のサーバシステムにおいて活用することを検討している。

今後の課題としては、定常的な運用を続けノウハウの蓄積をすることが挙げられる。

### 参考文献

- [1] 榎田 秀夫, 村田 和義, 渋谷 雄: 「京都工芸繊維大学における 10Gbps ネットワークインフラの導入について」, 情報処理学会 IOT 研究会, 第 2 回研究会, IOT6-4, June 27, 2009.
- [2] H.Masuda, K.Murata and Y.Shibuya: "Low TCO and High-Speed Network Infrastructure with Virtual Technology", ACM SIGUCCS 2009 Fall Conference, pp.321-324, October 11-15, 2009.
- [3] H.Masuda, K.Murata and Y.Shibuya and Y.Kuroe:

“High-Speed Network Infrastructure between KIT’s Campuses for Computer System Redundancy”, ACM SIGUCCS 2013 Service & Support Conference, pp.109-110, November 05-08, 2013.

- [4] eduroam JP <http://www.eduroam.jp/>.
- [5] Brocade Networks <http://www.brocadejapan.com/>.
- [6] Brocade Networks: Brocade VCS Fabric Technology, <http://www.brocade.com/solutions-technology/technology/vcs-technology/index.page>.
- [7] Fortinet, Inc: Fortigate 1000C, <http://www.fortinet.com/products/fortigate/1000C.html>.
- [8] Alaxala: AX3600S, <http://www.alaxala.com/jp/products/AX3600S/>.
- [9] Cisco AIR-CT5508, [http://www.cisco.com/web/JP/product/hs/wireless/5500wc/prodlit/data\\_sheet\\_c78-521631.html](http://www.cisco.com/web/JP/product/hs/wireless/5500wc/prodlit/data_sheet_c78-521631.html).
- [10] Cisco Virtual WLC, <http://www.cisco.com/web/JP/product/hs/wireless/vwc/index.html>.