

SPRING-8 および SACL A 制御・実験ネットワークの現状

STATUS OF CONTROL AND EXPERIMENTAL NETWORK SYSTEMS AT SPRING-8 AND SACL A

杉本 崇^{*A)}、石井 美保^{A)}、岡田 謙介^{A)}、大端 通^{A)}、坂本 達亮^{A)}、山鹿 光裕^{A)}、田中 良太郎^{A)}
Takashi Sugimoto^{*A)}, Miho Ishii^{A)}, Kensuke Okada^{A)}, Toru Ohata^{A)}, Tatsuaki Sakamoto^{A)},
Mitsuhiro Yamaga^{A)}, Ryotaro Tanaka^{A)}

^{A)}Japan Synchrotron Radiation Research Institute (JASRI/SPRING-8), 1-1-1 Kouto, Sayo, Hyogo, 679-5198, Japan.

Abstract

We report current status of network system at SPRING-8 and SACL A. The network of SPRING-8 control system have been remarkably changed during its 17-year-long operation. The X-ray free electron laser facility, SACL A, have just established in 2011. The number of network instruments of two facilities are reached more than 600. In addition, large bandwidth network are required to record image data at SACL A experiments. To maintain network systems of such large accelerator facilities by a few staff, we defined systematic rule of networking. Overview of the rules of networking at SPRING-8 and SACL A are described.

1. SPRING-8 および SACL A の制御システム

大型放射光施設 SPRING-8 は 1997 年よりコミッションおよび供用運転が開始され、現在に至るまで安定して運用されている。SPRING-8 の加速器およびビームラインの機器は MADOCA フレームワーク^[1]により制御されている。MADOCA では、制御室のワークステーション・データベースと現場の計算機・機器間の通信が、TCP/IP ネットワーク上のメッセージ交換により実現される。MADOCA の開発により、旧来のハードワイヤー制御から一線を画した、パケット通信ネットワークを前提とした制御システムを実現した。

2012 年 3 月より供用開始した X 線自由電子レーザー (XFEL) 施設 SACL A^[2]においても MADOCA が採用され^[3]、開発期間の短縮と高い信頼性を実現している。MADOCA は SACL A 加速器制御だけでなく実験計測制御にも採用され^[4]、フレームワークの活用範囲を広げている。

以上のように MADOCA フレームワークは、SPRING-8、SACL A の加速器制御システム、実験制御システムにおいて運用されている。MADOCA を支える計算機システム・ネットワークシステムの安定運用および性能も重要となる。本稿では、SPRING-8 および SACL A の制御システムを支えるネットワークシステムの現状について報告する。節 2. では論理ネットワーク構成について、節 3. では物理ネットワーク構成について説明する。

2. 論理ネットワーク構成

広義の「SPRING-8 サイト」は、兵庫県西播磨に位置する放射光実験研究施設である。SPRING-8 サイト内には、8 GeV 放射光リング「SPRING-8」とその入射器である線型加速器とブースターシンクロトロン、SPRING-8 と入射器を共用する兵庫県立大学 1.5 GeV 放射光リング「ニュースバル」、および、XFEL 施設「SACL A」が存在する。要求される帯域・セキュリティレベルに応じた機能分離により、オフィス系ネットワークを含め SPRING-8 サイト内には、約 870 台のネットワークスイッチ、約

410 台の無線 LAN、8 系統のファイアウォールが運用されている。^[5] 各施設の制御・実験ネットワークは、統一したポリシーで運用することによりセキュリティを担保しながら省力化・低コスト化を実現している。Table 1 に SPRING-8 サイト内の制御・実験ネットワークを示す。歴史的経緯のある一部セグメントを除き、2009 年に実施した SPRING-8 制御ネットワーク構成変更^[6]以降、一貫したポリシーで構築・運用している。

2.1 IP アドレス範囲

ネットワーク構成変更以降に整備した制御系ネットワークは第 2 オクテット (oct.) で区別し、各用途に/16 のアドレス空間を割り当てた。理論上 65,536 アドレスを使用可能であり、SPRING-8、SACL A 規模の加速器では十分なアドレス空間の広さである。用途単位を“LAN”と定義し、たとえば「SP8 制御 LAN」「XFEL DAQ-LAN」など呼称する。アクセスコントロールのために LAN 間をファイアウォールで隔離し、不用意に異なる LAN の機器にアクセスしないよう制限している。

ネットワーク構成変更を実施した理由の一つは、使用可能な IP アドレス数が逼迫していたことが挙げられる。ネットワーク構成変更以前は加速器制御が/21 (2,048 アドレス)のレイヤー 2(L2)接続の単一セグメントで構成され、前後アドレス領域の使用状況から拡張不可能な構成であった。別セグメントとしてビームライン制御が/23×4 (2,048 アドレス)、SCSS 試験加速器制御が/24 (256 アドレス)、が定義され、加速器制御と合計 4,352 アドレスである。SPRING-8 建設当時としては十分なアドレス空間であったが、デジタルマルチメーター等の Ethernet 接続機器の増加や、仮想マシン・コンテナ技術の適用^{[7][8]}により実計算機数に対する IP アドレスの使用量が著しく増加し、数年内に枯渇する見込みであった。

2.2 セグメント

ネットワーク構成変更後に定義したセグメントはサブネットマスクを/23 (512 アドレス)に、ゲートウェイアドレスをブロードキャストの一つ前に統一した。構成変更前までは、セグメントによって異なるサブネットマスク (/21, /23, /24)、一貫性のないゲートウェイアドレス

* takashi.sugimoto@spring8.or.jp

Table 1: Network systems for accelerator controls and experiments in the SPring-8 site. We classified network system by purpose. This table shows DNS subdomain, IP address range, and subnet mask of each LAN. For access control, we segregate each LAN using firewalls. (*A few network segment in the SPring-8 are not standardized due to historical reason.)

用途	LAN 通称	サブドメイン	IP アドレス範囲	サブネットマスク
SPring-8・ニュースバル制御	SP8 制御	sp8.cntl	172.20.0.0/16 172.24.0.0/16	/23 /23, /24*
SPring-8 利用実験・実験計測	BL-USER	blu.sp8.cntl	192.168.0.0/16	/24*
SACLA 加速器制御	XFEL 制御	xfel.cntl	172.16.0.0/16	/23
SACLA 利用実験	XFEL USER	user.xfel.cntl	172.30.0.0/16	/23
SACLA 実験計測	XFEL DAQ	daq.xfel.cntl	172.28.0.0/16	/23
プログラム開発系	中立開発	dmz.cntl	172.26.0.0/16	/23

を使用していたため、設定ミスによる通信トラブルが少なからず発生していた。セグメント設計の統一により、機器追加・入れ替え時のネットワーク設定ミスに起因するトラブルを減らすことができた。

1セグメントのIPアドレス空間を512にした理由は、1セグメントの範囲が広すぎると増加したブロードキャストパケットにより一部の組み込み機器が誤動作するためである。^[9] ネットワーク構成変更以降、頻繁に誤動作を起こしていたモーターコントロールユニット、デジタルマルチメーター、オシロスコープ等の障害がほぼゼロとなった。

2.3 名前解決とドメイン名

ネットワーク構成変更と同時に、名前解決を Network Information Service (NIS) から、Domain Name System (DNS) に移行した。NIS はスケラビリティに乏しく、SPring-8、SACLA、および開発系それぞれで NIS ホストマップを分散管理し、お互いに名前解決できる構成をとることが困難であった。また、HP-UX、Solaris、Linux 間における NIS 実装の非互換性や、NIS をサポートしない Windows 端末の増加に伴い、名前解決の標準規格である DNS の導入が必須な状況であった。

DNS の導入にあたり、名前空間（ドメイン名）を設計した。第一サブドメインとして制御システムを表す“cntl”、第二サブドメインとして加速器施設種別を表す“sp8”、“xfel”、および開発系として“dmz”を定義した。さらに第三サブドメインに SPring-8 利用実験・実験計測を表す“blu”、SACLA 利用実験を表す“user”、SACLA 実験計測“daq”を定義した。ドメイン名からオフィスネットワークと制御システムを明確に区別し、SPring-8、SACLA とその用途を判断できるようになった。一方、デメリットとして Fully Qualified Domain Name (FQDN) が長くなる。しかしながら一般的なオペレーティングシステムは優先検索ドメインを指定可能であり、同一 LAN 内では従来通りホスト名のみで通信可能である。

ドメイン空間を分けたもう一つの理由は、SPring-8 制御系と SACLA 制御系の弱結合の実現である。通常の加速器運転時は別個の制御系として運用し、ある制御系の操作が他の制御系に影響することは固く禁じる必要がある。一方で、近い将来予定されている SACLA 線型加速器から SPring-8 蓄積リングへの電子ビーム入射運転では、SACLA と SPring-8 制御系が連携動作できる必要がある。通常運転時は優先検索ドメインを利用するこ

とでホスト名のみで通信し、連携動作時は GUI・アプリケーション等で FQDN を指定することにより異なる LAN と通信していることを操作者に明確に意識させることができる。

2.4 DNS, NTP 等のネットワークサービス

DNS や、Network Time Protocol (NTP) による時刻同期、Simple Mail Transfer Protocol (SMTP) によるメールリレーなど、計算機運用上必須なサービスも系統的に整備した。キャッシュDNS サーバーは、各 LAN の第3、第4 oct. が 2.1 および 3.1（例えば、XFEL 制御 LAN では 172.16.2.1, 172.16.3.1）に統一した。SPring-8、SACLA の制御系や実験計測系の複数 LAN をまたいだ機器担当者は、「DNS 設定は必ず*.2.1, *.3.1 である」という単一のルールを知っていればよい。

NTP や SMTP サーバーも同様に、全ての LAN において単一ルールで整備した。NTP サーバーのホスト名は ntp1, ntp2, ntp3、SMTP サーバーのホスト名は mail1, mail2 である。NTP や SMTP が一つしか指定できないアプライアンス・組み込み機器は数字指定なしの ntp, mail を使用し、DNS ラウンドロビンによりいずれかの実サーバーに接続される。

3. 物理ネットワーク構成

SPring-8 および SACLA の各 LAN は、レイヤー 3 スイッチ (L3SW) を頂点としたスター構成のスイッチド・ネットワークである。ネットワークスイッチの機器数は SPring-8 全体で約 400 台、SACLA 全体で約 270 台である。機器数が多い理由は、銅線ケーブルによるグラウンドループの形成を避けるため、ほぼ全てのラックに光接続のネットワークスイッチを実装しているためである。

3.1 ネットワークスイッチに対する性能要件

物理構成はスター型として一般的な 3-Tier Network であり、コア層（多ポート光接続 L3SW）、分配層（多ポート光接続 L2SW）、およびアクセス層（アップリンク光接続の L2SW）で構成される。各 LAN ともコア層・分配層に信頼性の高い機種を採用するとともに、数量が多いアクセス層を安価な機種とすることで、低コスト化を実現している。

Table 2 に、SPring-8、SACLA で運用中のネットワーク機器を用途、性能等の要件で分類した。2014 年時点では、すべての機器を買い取りで保有・運用している。

Table 2: Network instruments installed in the SPring-8 and SACLA control systems and experiments. The network instruments are classified by tiers, port throughput, and other requirements.

用途	uplink	downlink	その他の要件	運用数	主な機種
ファイアウォール (Firewall)	-	1G		3組	Cisco ASA5525-X, 5545-X
	-	1G	アプリケーション識別	1組	PaloAlto PA-2050
コア層 (Core)	-	10G	L3, スタック, 分散 LAG	3組	HP A5900, A5820
	-	1G	L3, スタック, 分散 LAG	2組	Cisco Catalyst 6500-E, 3750-E
分配層 (Distribution)	10G	10G	スタック, 分散 LAG	18	HP A5820
	1G	1G	スタック or 冗長電源	41	HP A5500EI, 5400zl
アクセス層 (Access)	10G	10G	計算機室, 冗長電源	5	HP A5820AF
	10G	1G	計算機室, 冗長電源	15	HP A5830AF
	1G	1G	計算機室, 冗長電源	25	HP 6600
	10G	1G		22	HP A5800, 2910
	1G	1G		70	HP A5120EI, 2910, 2810, 2530
	1G	1G	PoE+	18	HP 2910, 2530
	1G	100M		223	HP 2600, 2610, 2510
	1G	100M	PoE/PoE+	221	HP 2600, 2610, 2620, MNO M24

多数の機器のうち、ファイアウォールとコア層のみ保守契約を結んでいる。分配層以下は要件分類に応じた同等予備品を少数保有することで故障対応している。予備品はオフィスネットワーク用と共通して整備しており、各用途ごとに1台以上かつ運用機器数の5%以下を目処として保有している。

コストを削減する鍵として、ライフタイム保証製品の採用がある。Hewlett Packard (HP) 社製品の大半はライフタイム保証による長期にわたる故障交換を受けられるため、追加費用無しで予備品数を維持することができる。近年はCisco, DELL等の他社もライフタイム保証を提供しているため、今後も要件に応じた最適機種を選定していく予定である。

3.2 物理ネットワークと論理ネットワークの対応関係

SPring-8, SACLAのネットワークシステムは、Virtual LAN (IEEE802.1Q, VLAN) を用い、セグメント分割している。レイヤー2であるVLANとレイヤー3のセグメント(ネットワークアドレス)を以下の式により系統的に対応づけた。

$$(\text{VLAN ID}) = (\text{第2 oct.}) \times 100 + (\text{第3 oct.}) \quad (1)$$

例えば、XFEL制御LANのDNSサーバー(172.16.2.1, 172.16.3.1)が収容されるセグメント(172.16.2.0/23)のVLAN IDは、式1より1602である。

VLAN IDとセグメントの系統性により人間によるネットワーク設定作業が直感的になるとともに、運用監視スクリプト等が外部テーブル等の参照無しに実行可能となった。

3.3 広帯域実験データ転送

SACLAにおけるコヒーレントX線回折イメージング^[10]は、特に大量のデータ転送が見込まれる実験の一つである。SACLAで供用中のMPCCD検出器^[11]は、12センサーの最大構成時に6Gbpsのデータレートに達する。さらに2014年冬以降、SACLAビームライン

3 (BL3)とBL2の同時振り分け実験が予定されている。BL3とBL2で同時にデータ収集する場合はデータレートが12Gbpsに達し、10Gigabit Ethernet (10GbE)では帯域が不足する。ビームライン同時実験における大量のデータ転送に対応するため、XFEL DAQ-LANの物理層は3-Tier Networkモデルを拡張して実装した。

Figure 1は、XFEL DAQ-LANの概念図である。MPCCDの画像データは分配層-コア層-分配層間の帯域共有区間を経由せず、キャッシュストレージのアクセス層ネットワークスイッチに直接転送される。その他のビームラインモニター類は合計1Gbps程度であるため、分配層・コア層を経由して転送される。以上のように、アクセス層スイッチ間の直接接続を整備することにより、まだ高価な40GbE, 100GbEを採用することなく複数ビームライン同時振り分け実験を成立させる。

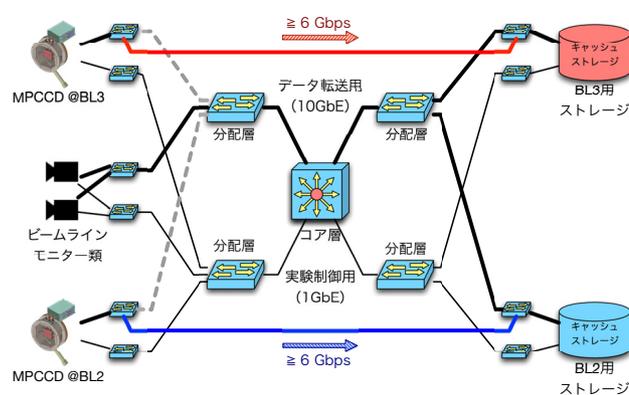


Figure 1: Schematic diagram of physical layer of XFEL DAQ-LAN. Left side and right side represents experimental hall and computer room of SACLA, respectively. Total data rate per beamline is more than 6 Gbps. To perform experiments at beamline 2 and 3 simultaneously, we design special network route from access switch of MPCCD to access switch of cache storage. Other image data, such as beamline monitors (~ 1 Gbps), are transferred using general network route.

実験データは SPring-8 サイト内に整備した 13 TFlops x86 クラスター計算機、90 TFlops スーパーコンピューター FX10、あるいは、サイト外のスーパーコンピューターに転送され解析する。特に「京」への 5 Gbps 超高速データ転送を目標として転送フレームワークの構築と、ネットワークのチューニングを現在進めている。^[12]

4. まとめ

2009 年以降、SPring-8 および SACLA の制御系・実験系ネットワークを統一したポリシーで整備している。2014 年現在、オフィスネットワークを含めて約 870 台のネットワーク機器を数名のスタッフで管理している。ポリシーの統一により論理・物理構成がシンプルになり、ネットワーク障害発生を減らすことに成功した。今後、さらにデータ量が増える見込みの SACLA 利用実験への対応を強化しつつ、既存構成と整合性のあるポリシー拡張を行っていく予定である。

参考文献

- [1] R. Tanaka, et al., “The first operation of control system at the SPring-8 storage ring”, The 6th International Conference on Accelerator and Large Experimental Physics Control Systems (ICALEPCS'97), Beijing, China, 1997.
- [2] T. Ishikawa et al., “A compact X-ray free-electron laser emitting in the sub-angstrom region.”, *Nature Photonics* 6, 540 (2012).
- [3] 福井達, ほか, 「SACLA 加速器制御システム」, 放射光 Vol.25, No.2, pp.65-69 (2012).
- [4] M. Yamaga, et al., “Data Acquisition System of over Giga-Bps of Data Rate for User Experiment at X-Ray Free-Electron Laser Facility SACLA”, The 19th REAL-TIME Conference (RT2014), Nara, JAPAN, May 26–30, 2014.
- [5] 杉本崇, 「SPring-8 ネットワークシステムの現状」, 第 17 回共同利用機関におけるセキュリティーワークショップ (SWS2013), 国立天文台水沢キャンパス 奥州宇宙遊学館, 2013 年 10 月 24–25 日.
- [6] T. Sugimoto, M. Ishii, T. Ohata, T. Sakamoto, and R. Tanaka, “Reassignment of network addresses at SPring-8 control system”, The 7th International Workshop on Accelerator Operations (WAO2010), INTEC, KAERI, Daejeon, KOREA, April 12–16, 2010.
- [7] M. Ishii, T. Fukui, T. Masuda, T. Nakatani, T. Ohata, “A Software Framework to control a network-connected equipment as a pseudo device”, The 9th International Conference on Accelerator and Large Experimental Physics Control Systems (ICALEPCS2003), Gyeongju, Korea, 2003.
- [8] T. Masuda, T. Ohata, T. Fukui, “Application of a Virtualization Technology to VME Controllers”, The 11th International Conference on Accelerator and Large Experimental Physics Control Systems (ICALEPCS2007), Knoxville Convention Center, Knoxville, Tennessee, USA, October 15–19, 2007.
- [9] T. Sugimoto, et al., “A study of network vulnerability in embedded devices”, The 2nd Control System Cyber-security Workshop, Kobe, Japan, October 2009.
- [10] K. J. Gaffney and H. N. Chapman, “Imaging atomic structure and dynamics with ultrafast X-ray scattering,” *Science* **316**, 1444 (2007).
- [11] T. Kameshima et al., “Development status of X-ray 2D detectors for SPring-8 XFEL,” IEEE Nuclear Science Symposium, Knoxville, TN, USA, October 2010.
- [12] K. Okada, et al., “Data Acquisition Framework at SACLA”, IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Coex, Seoul, KOREA, October 27 – November 2, 2013.