

Title: NSPIXP Working Group Activities in 2003

Author(s):

加藤 朗 (kato@wide.ad.jp)
関谷 勇司 (sekiya@wide.ad.jp)
中村 修 (osamu@wide.ad.jp)
重近 範行 (nazo@wide.ad.jp)
山口 英 (suguru@is.naist.jp)
森島 直人 (naoto@dl.naist.jp)
宇多 仁 (zin@jaist.ac.jp)
長谷部 克幸 (hasebe@wide.ad.jp)
安田 歩 (yasuda@ocn.ad.jp)
岡本 裕子 (yuko@mcnet.ad.jp)

Date: 2004 年 7 月 25 日

目次

1	はじめに	2
2	DIX-IE (Distributed IX In EDO)	2
2.1	接続組織数	2
2.2	運用の履歴	3
2.3	トラフィックの推移	5
3	NSPIXP-3	8
3.1	接続組織数	9
3.2	運用の履歴	9
3.3	トラフィックの推移	9
4	関連イベント	10
5	他の IX 運用者との情報交換	13
	wide-draft-nspixp-activity-2003-03.pdf	1

1 はじめに

NSPIXP (Network Service Provider Internet eXchange Project) は、商用インターネットを相互に接続する場合の問題点について、実証的な手法で研究をすすめるプロジェクトである。このドキュメントでは、NSPIXP Working Group が 2003 年におこなった活動内容や、NSPIXP の現状について記す。

2 DIX-IE (Distributed IX In EDO)

NSPIXP では、かねてより構成拠点の分散による分散 IX の可能性を検討してきた。KDDI 大手町ビルに設置された NSPIXP-2 の分散化に関しては、2000 年度の WIDE 報告書においてその方向性を示し、2001 年度および 2002 年度の報告書において実験内容及びその構成を示した。

そして、2003 年 3 月 28 日、ISP の接続ポイントを都内複数拠点に分散した NSPIXP-2 が実運用にはいった。また、同時に名称を DIX-IE(Distributed IX in EDO) へと変更し、IPv6 へも対応した。現在、DIX-IE は以下の 6 拠点で構成されている。

- KDDI(WIDE Project) 大手町
- MCI 新川
- MIND 西大井
- NTT Communications 大手町
- @Tokyo 豊洲
- AboveNet 日本橋

上記の拠点間は、10 Gigabit Ethernet や Gigabit Ethernet のリンクアグリゲーションをもちいることにより、4~20Gbps の帯域で相互接続されている。現在の構成図を図 1 に示す。

2.1 接続組織数

2004 年 1 月 4 日現在、DIX-IE に接続している組織数は 73 である。また、1 組織がトラフィック増加への対応や冗長性確保のために、複数ポートで接続するケースもめずらしくない。そのため、接続ポート数は接続組織数を大きく上回る 104 ポートにのぼる。これらをメディアの種別ごとに整理すると、以下のようになる。

- 10 Gigabit Ethernet 1
- Gigabit Ethernet 80

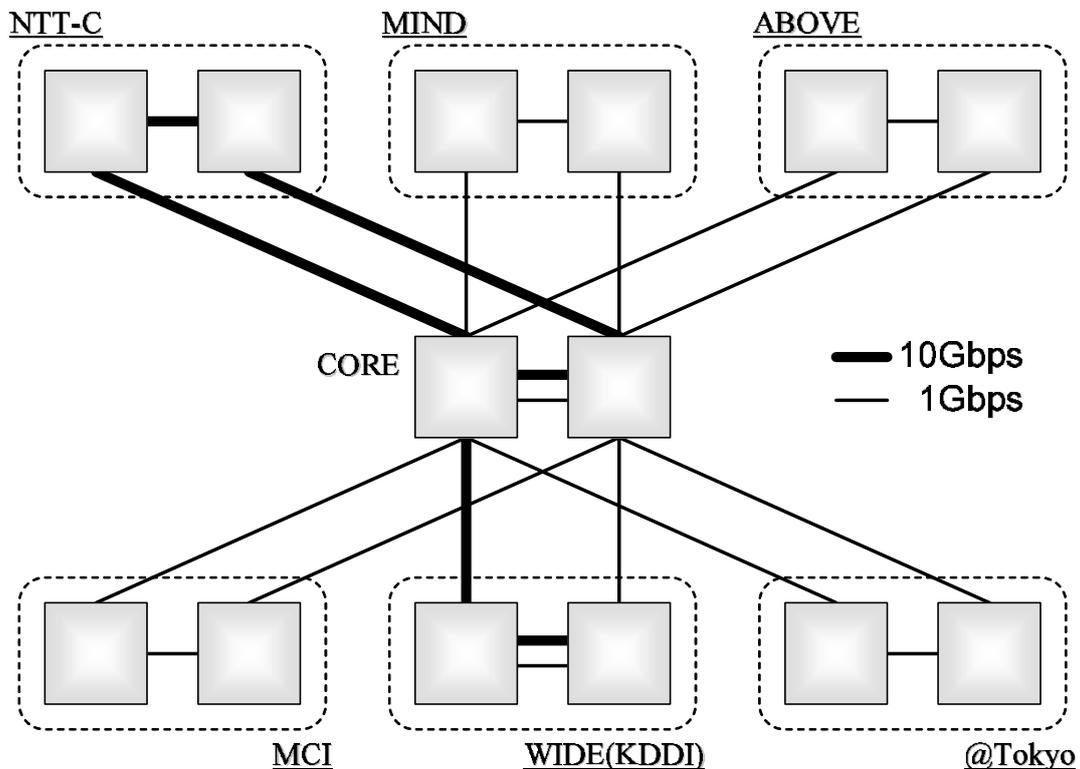


図 1: 現在の DIX-IE の構成図

- Fast Ethernet 23

トラフィック増加への対応に、10 Gigabit Ethernet での接続を検討している組織が徐々に現れはじめており、上記のようにすでに接続している組織も現れはじめています。また、接続ポート数を分散拠点毎に整理したものは以下のとおり。

- KDDI(WIDE Project) 74
- MCI (拠点移設中) 2
- MIND 4
- NTT Communications 18
- @Tokyo 2
- AboveNet 6

DIX-IE の運用は開始されて日が浅いため、NSPIX-2 の設置拠点であった KDDI(WIDE Project) が圧倒的に多いことがわかる。

2.2 運用の履歴

冒頭で述べたように、2001 年 6 月より検討してきた分散 NSPIX-2 は、2003 年 3 月 28 日より DIX-IE として実運用にはいった。実際には、かねてから問題点の洗い出し

などのために試験運用をつづけてきた分散 NSPIXP-2 を、2003 年 3 月 24 日に既存の NSPIXP-2 と接続し、移行準備を開始した。

2.2.1 監視体制および監視対象

監視体制としては、各拠点運用チームによる 24 時間のスイッチ状態監視と WIDE Project NSPIXP メンバによる全体監視の 2 通りによる体制をとっている。監視方法としては、各拠点スイッチ間のトラフィックを交換する回線と別回線をもちいる OutBound Ethernet 網の 2 つの segment を用いることで、各拠点スイッチ全体の状態監視と、DIX-IE の拠点間状況を把握することができる構成となっている。

また、各拠点のトラフィックを交換する回線上に、監視対象装置を設置し、KDDI 大手町に設置した端末から各拠点に設置した監視対象装置の監視を行なうこととした。

2.2.2 10 Gigabit Ethernet での障害

2003 年 6 月より、10 Gigabit Ethernet をもちいた拠点間および拠点内接続の一部で、データ損失を含む障害が発生した。これはラインカード個体の受光許容範囲の問題であると思われ、ラインカードの交換などの緊急メンテナンスを実施した。

2.2.3 ファームウェア更新

VPNなどを主な用途として MPLS(Multi-Protocol Label Switching)の需要が高まっている。MPLSでは、データリンク層ヘッダとIP層ヘッダの間にラベルと呼ばれるデータを挿入する。したがって、経路上の機器は1500バイトを越えるMTU、いわゆるJumbo Frameに対応しなければならない。

近年、通信事業者間がMPLSで接続されることも多くなっており、DIX-IEでもJumbo Frameへの対応が必要となった。このことから、Jumbo Frameへ対応するために、2003年10月30日、KDDI(WIDE Project)拠点のスイッチにおいて、ファームの更新作業をおこなった。

なお、このファームウェアでは、同時にBPDU TCNの不具合修正も兼ねている。

2.2.4 IEEE802.1w Spanning-Tree Parameter

ファームウェアの交換にともない、2003年10月30日にKDDI(WIDE Project)拠点のスイッチにおいて、新たに定義された

$$\begin{aligned} \text{max-age} &\geq \text{forward-delay} \\ (2 * \text{forward-delay} - 1) &\geq \text{max-age} \geq (2 * \text{hello-interval} + 1) \end{aligned}$$

の条件を満たすため、IEEE 802.1wのパラメータを $\text{forward-delay} = 9 / \text{hello-interval} = 2 / \text{max-age} = 9$ から $\text{forward-delay} = 9 / \text{hello-interval} = 1 / \text{max-age} = 10$ に変更し

た。各拠点のスイッチにおいては 2003 年 12 月 2 日に実施した。

2.2.5 未解決の問題

2003 年 12 月 2 日 19 時 5 分頃、KDDI-Above 拠点間において通信断をともなう障害が発生した。これは、

- Spanning-Tree(IEEE802.1w) の状態 Forwarding
- ポートの状態 Down
- リンクアグリゲーション (IEEE802.3ad) InActive

という状態になり、通信が全く不可能な状態になるものである。BPDU の Hello がタイムアウトした時点で発生していると推測されているが、完全には原因が解明されておらず、現在も調査中である。なお、この障害は上記の時間帯のほかに、さらに一度、異なるスイッチおよびインタフェイス間でも発生している。

2.3 トラフィックの推移

本節では、DIX-IE で交換されているトラフィックに関して述べる。

図 2 のグラフは、NSPIXP-2 の稼働時からの総トラフィック量の推移を示したものである。ここで総トラヒックとは、ISP を接続しているポートの流入トラフィックの総和である。2003 年 11 月後半にみられるトラフィックの減少傾向は、P2P アプリケーションを利用して著作権物の違法交換をしていたものが、逮捕された影響であると推測される。

図 3 は、図 2 の Y 軸のみをログスケールに変換したものである。緑色で示した直線は

$$Traffic = b * \exp(a * year)$$

で示される近似であり、 a は 0.657 である。この図の通り、DIX-IE で交換されているトラフィックはこの近似線に非常によく沿った形での増加傾向を見せている。この増加傾向はおおよそ毎年 2 倍であり、この傾向がこのまま継続すると仮定すると、2004 年末には 30Gbps に達すると予想される。

図 4 は、2002 年 11 月～2003 年 1 月のトラフィックを 1 日単位で示したものである。ISDN がインターネットへ接続するための主な手段であった数年前は、NTT のテレホーダイサービスの影響で、午後 11 時にトラフィックの急激な増加が観測された。現在では、ケーブルテレビや ADSL、あるいは FTTH のような、いわゆるブロードバンド技術がインターネットへの主な接続手段となり、その結果として過去にみられたような増加はみられなくなった。

また、平日は午後 1 時前後に若干のトラフィックの増加が見られる。これは、昼食後の休みにホームページを閲覧するなどの習慣によるものであると思われる。

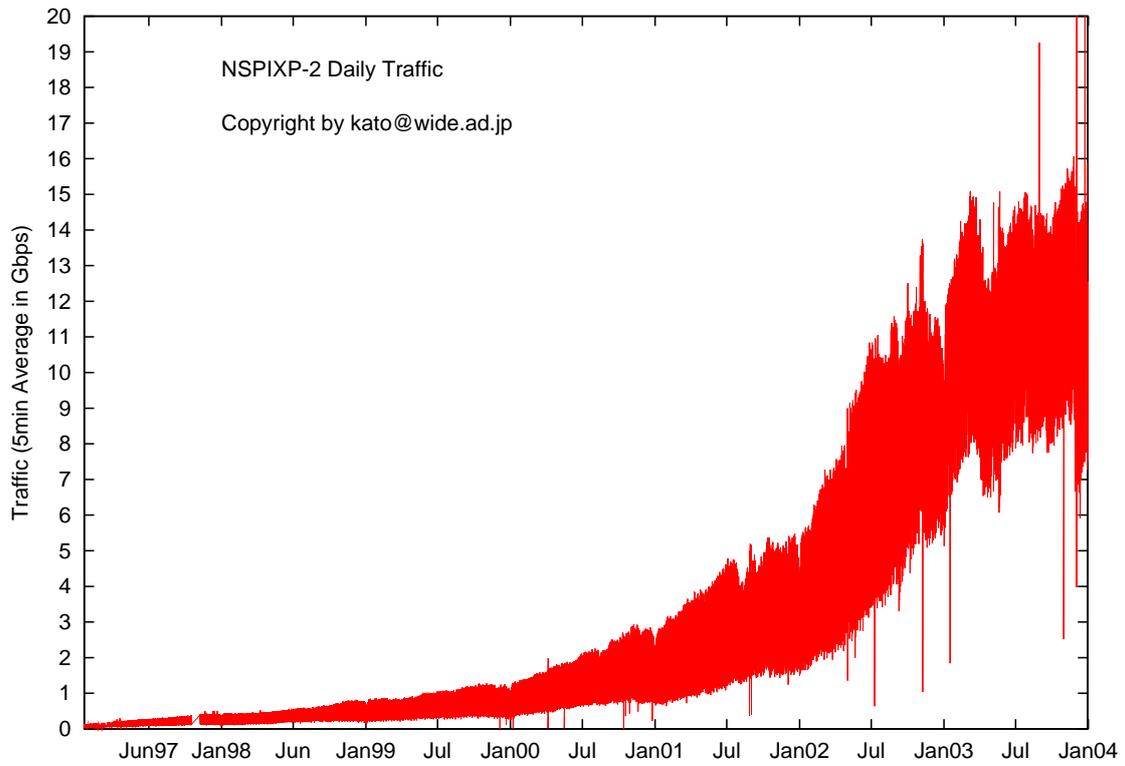


図 2: 現在までの総トラフィックの推移

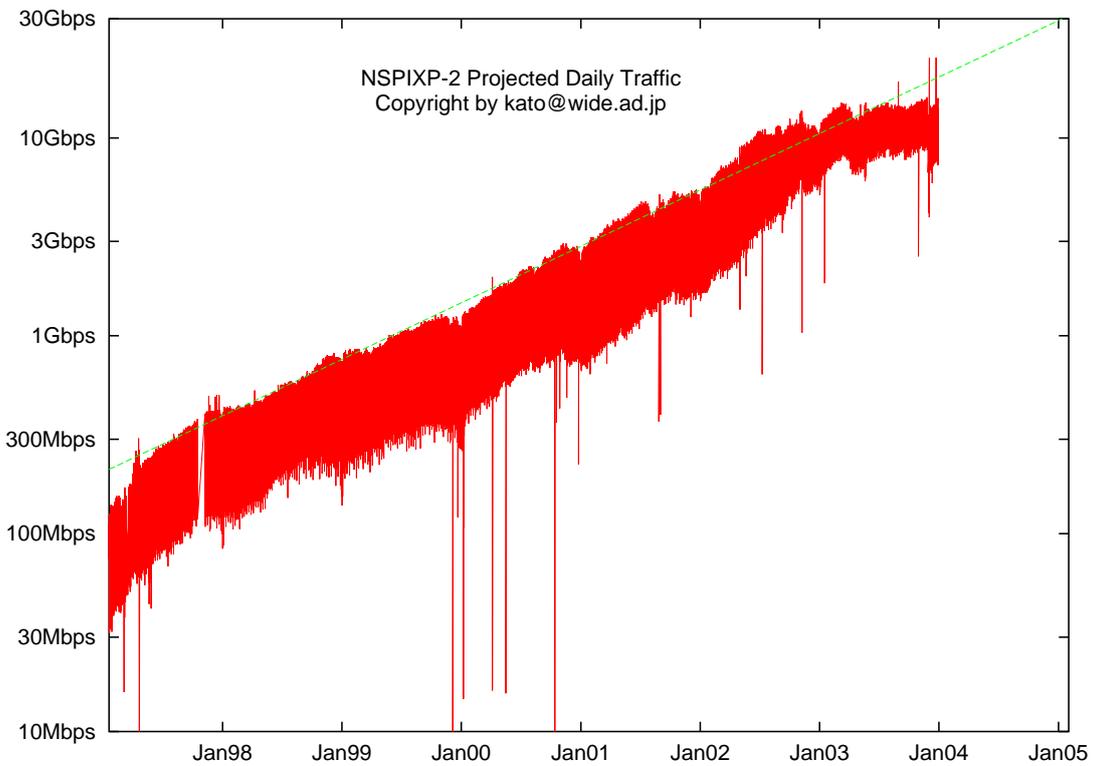


図 3: 現在までの総トラフィックの推移 (ログスケール)

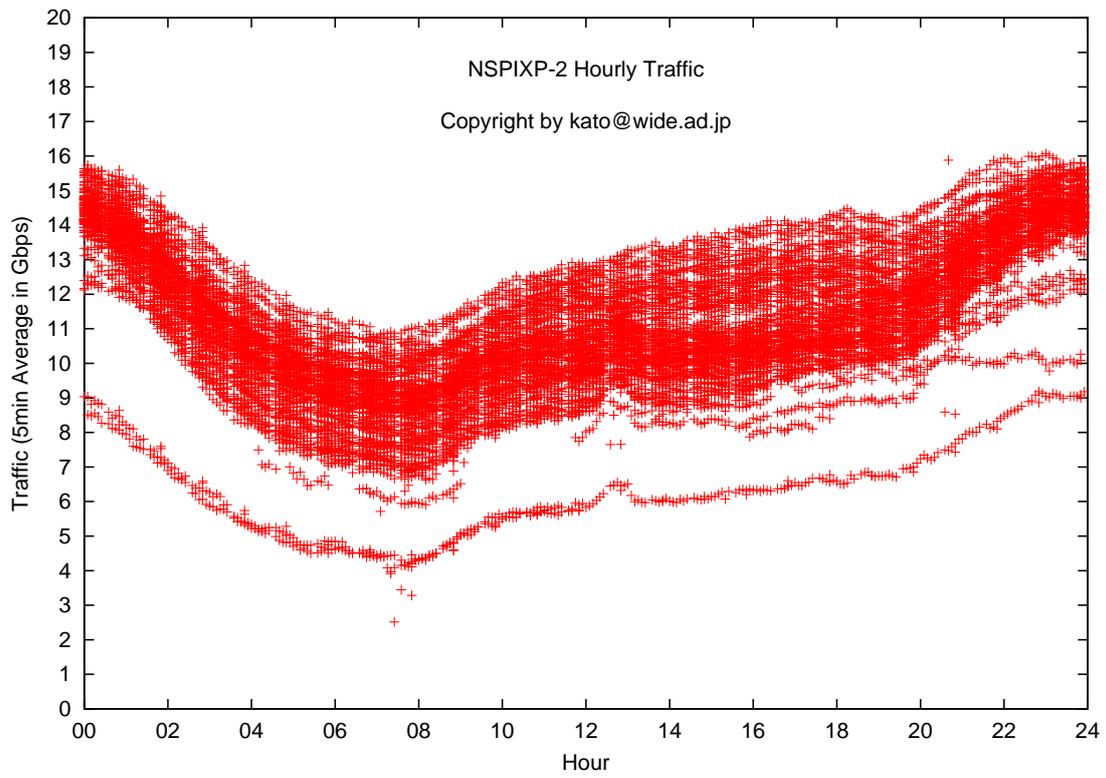


図 4: 一日のトラフィックの推移

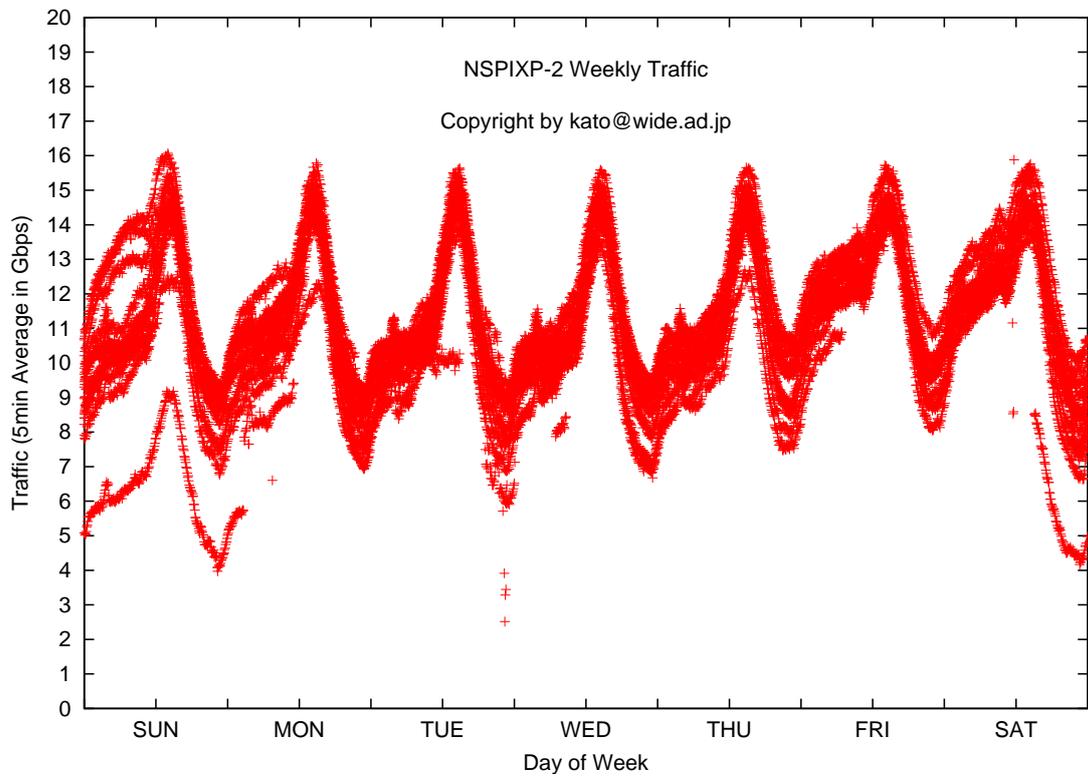


図 5: 一週間のトラフィックの推移

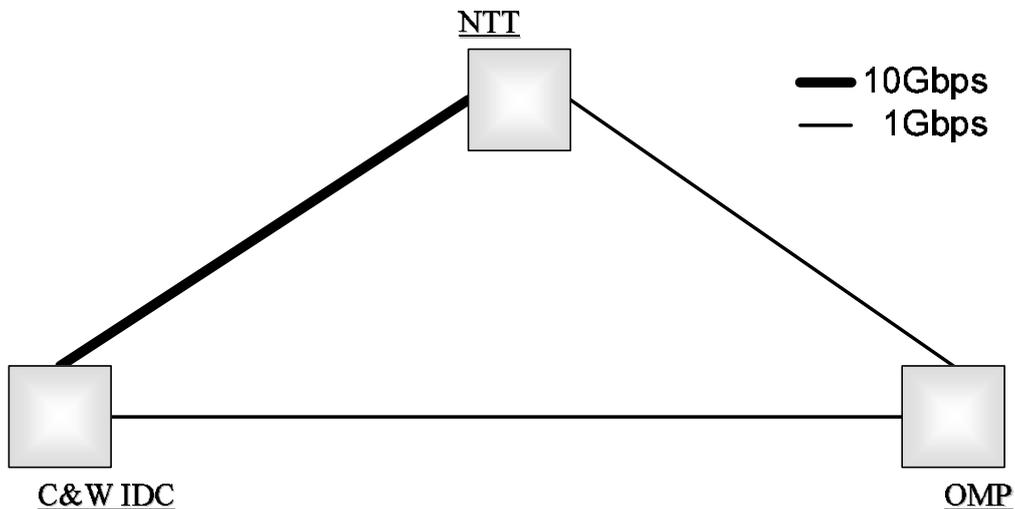


図 6: 現在の NSPIX-3 の構成図

図 5 は、図 4 のデータを 1 週間単位で示したものである。インターネットの黎明期には週末に大きな減少がみられたが、現在ではそのような減少は観測されず、平日と週末でほぼ同様のトラフィックがみられる。

3 NSPIX-3

NSPIX-3 は、1996 年に関西で初の IX として設置された。また、設置当初より分散 IX を意識しており、2 拠点の構成で運用が開始された。現在は拠点が増強され、以下の 3 拠点となっている。

- NTT 堂島
- C&W IDC 福島
- OMP 湊町

上記の 3 拠点はすべて相互に接続する構成になっており、トライアングルを形成している。なお、拠点間の接続は以下の通り。

- 堂島～福島 10 Gigabit Ethernet
- 福島～湊町 Gigabit Ethernet
- 湊町～堂島 Gigabit Ethernet

このように、拠点間は 10Gbps あるいは 1Gbps での接続となっている。また、トライアングルは IEEE802.1w Spanning-Tree によりループを回避しており、定常状態では湊町～堂島拠点間でブロックしている。現在の構成図を図 6 に示す。

3.1 接続組織数

2004年1月4日現在、NSPIXP-3に接続している組織数は24である。また、dix-ieと同様に、トラフィック増加への対応や冗長性確保のために複数ポートで接続している組織もあり、接続ポート数は25となっている。メディア種別ごとの接続数は以下の通り。

- Gigabit Ethernet 15
- Fast Ethernet 9

また、接続ポート数を拠点ごとに整理したものは以下の通り。

- NTT 16
- C&W IDC 6
- OMP 2

3.2 運用の履歴

NSPIXP-3では、2002年10月4日に現在の構成と旧構成を相互に接続し、新構成へと移行を開始した。また、この移行では拠点間リンクとして新たに10 Gigabit Ethernetが導入された。移行は2003年2月には一部を除きほぼ完了した。

3.2.1 ファームウェア更新

利用中のファームウェアにおいて、10 Gigabit Ethernetポートの入力カウンタが正常に動作しない不具合があったため、10月23日にすべての拠点のスイッチのファームウェアを更新した。この作業により、福島～堂島拠点間のトラフィックを入力側で計測することができなかった問題が解決された。

3.3 トラフィックの推移

本節では、NSPIXP-3で交換されているトラフィックに関して述べる。

図7のグラフは、NSPIXP-3で交換されている総トラフィック量の推移を示したものである。2001年7月～2002年10月までの取得データに関しては、縮退して保存する手法をとっていたため、1日平均の最大値のみとなっている。また、2002年10月～2003年3月までは、NSPIXP-3の構成変更にもなって技術的に正しい値が取得できていないため、表示していない。

この図にみるように、今年度のNSPIXP-3のトラフィックは著しい増加がみられた。2002年1月末のトラフィックを0.7Gbpsであると仮定すると、2003年1月末までの1

年間で約 5 倍増加したことになる。これは、接続 ISP が負荷分散や災害時などに備えた冗長化のために、東京の一極集中から地理的に分散した大阪へトラフィックを逃がし始めたためであるとみられている。

図 8 はのグラフは、図 7 の Y 軸のみをログスケールに変換したものである。緑色で示した直線は、

$$Traffic = b * \exp(a * epoch)$$

で示される近似である。ここで、 a は $3.92421e - 08$ 、 $epoch$ は 1970 年 1 月 1 日からの経過秒数である。DIX-IE と同様に、非常に近似に沿った形での増加をつづけており、2005 年末には 10Gbps に達すると予想されている。

図 9、および、図 10 のグラフは、2002 年 11 月～2003 年 1 月のトラフィックをそれぞれ 1 日単位と 1 週間単位で示したものである。グラフのばらつきからも、急激な伸びをみることができる。

4 関連イベント

NSPIXP では、一時的に接続性を必要とするイベントなど協力している。ここでは、毎年恒例になっている、NetWorld+Interop への接続性の提供に関して報告する。

毎年、6 月の終わりから 7 月のはじめにかけて、千葉県の幕張メッセにおいて Networld+Interop(以下、N+I) と呼ばれるイベントが開催されている。このイベントは、インターネットに関する技術の展覧会であり、各種のベンダがあたらしい技術を実装した機材の展示やデモンストレーションをおこなう。会場内には、出展社の機材を収容するためのネットワークが敷設される。このネットワークは、標準化される前の技術などを駆使したものであり、ここで利用された技術がつぎの年の主流になっているなど、インターネット技術の最先端を結集したものとなっている。

N+I では、各出展社や関係施設に対してインターネットへの接続性を提供する、ShowNet と呼ばれるネットワークが構築される。ShowNet は独自のアドレスブロック及び AS 番号をもち、全世界を開催される展示会とともに移動するイベントネットワークである。東京開催時には、東京の主要な IX に接続して各 ISP との間に peer/transit をおこなう。NSPIXP は ShowNet の対外接続において中心的役目を果たしている。

『NetWorld+Interop 2003 東京』は、2003 年 6 月 30 日から 7 月 4 日にかけて幕張メッセで開催された。NSPIXP との接続は、大手町に位置する NTT Communications 拠点でおこない、DIX-IE および NSPIXP-6 と接続した。DIX-IE に対しては、ShowNet のルータからの 10 Gigabit Ethernet LR による接続、および、L2 スイッチからの Gigabit Ethernet LX による接続の 2 種類のメディアを利用した。また、DIX-IE における ShowNet への Peer/Transit 組織数は 28 を数えた。一方、NSPIXP-6 に対しては 100Bae-Tx を利用し、Peer/Transit 組織数は 19 であった。

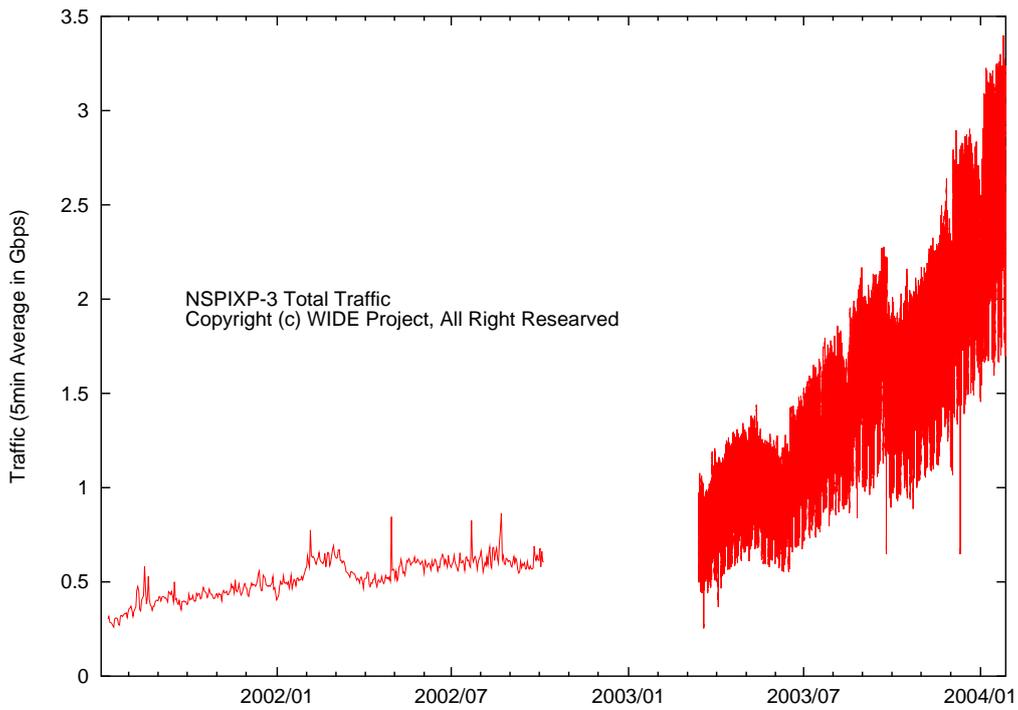


図 7: 現在までの総トラフィックの推移

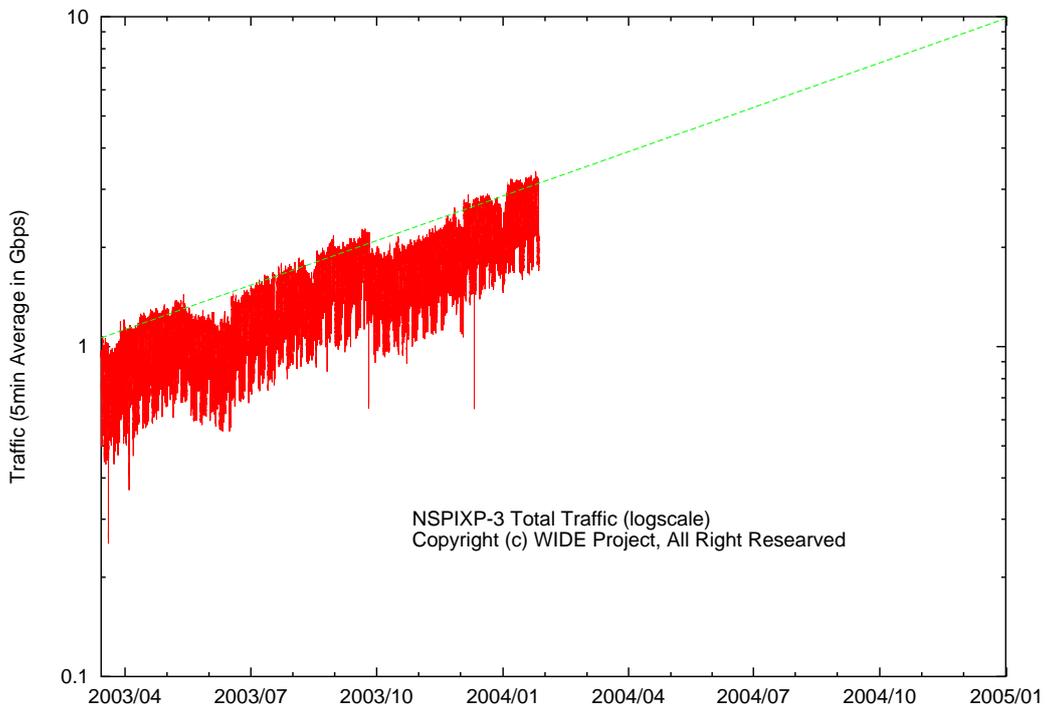


図 8: 現在までの総トラフィックの推移 (ログスケール)

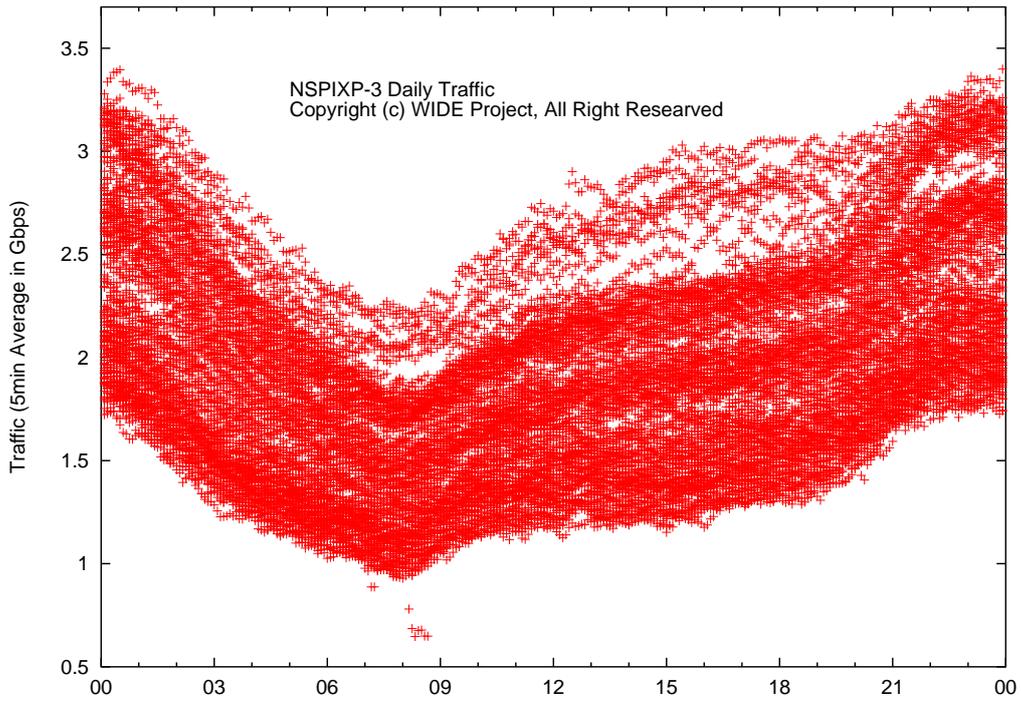


図 9: 一日のトラフィックの推移

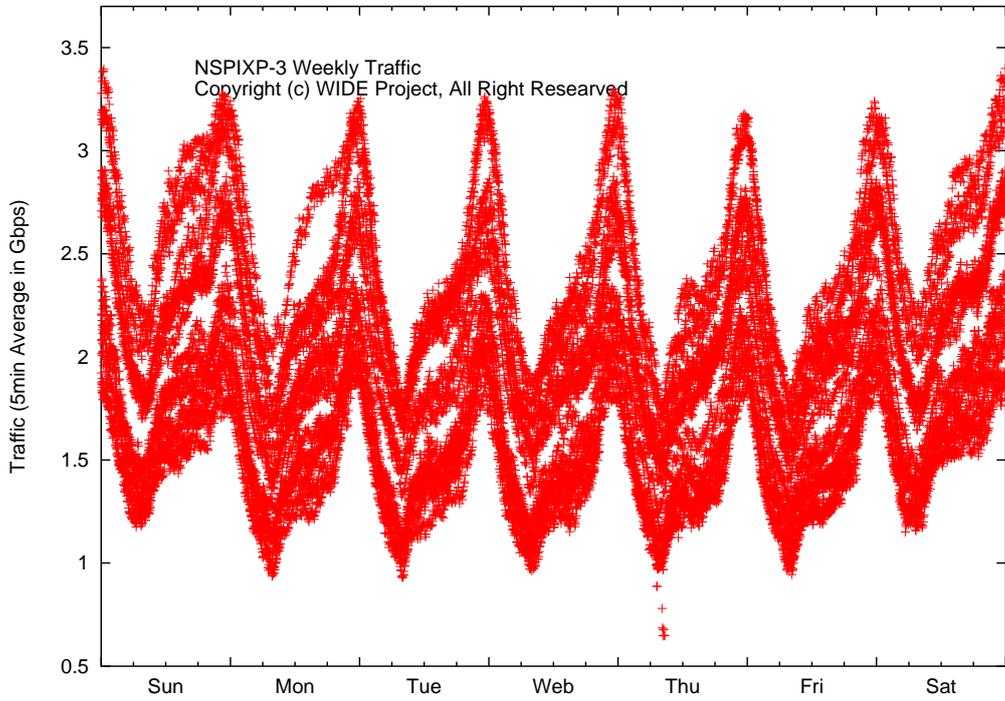


図 10: 一週間のトラフィックの推移

5 他の IX 運用者との情報交換

日本の主要な IX は、NSPIXP/JPIX/JPNAP がある。これらは一部で競合関係にある反面、障害発生時にトラフィックを相互に逃がすなどの協力が必須である。このようなことから、上記の主要 IX 間で情報交換をおこなうための連絡会である IX-exchanges を開催している。今年度は、2004 年 2 月 2 日に第 2 回会合が開催された。

第 2 回においては、近年のインターネットの急速な普及、アクセス回線の高速化および低価格化によるトラフィックの急増により、IX 間の情報交換だけでなく ISP を含めて連携する必要があるという観点から、ISP3 社を招いて議論が行われた。各 IX では、複数拠点での接続提供により traffic の分散が行われ、また ISP からは 10GE 接続要求が出始めるなど、アーキテクチャおよび効率を考えなければ行けない状況である。今後とも ISP のトラフィック動向や接続状況の情報交換を行い、各 IX においてどのようなトラフィック分散をするべきか、どのような構成で IX を形成するか議論が必要であり、継続することとなった。

Copyright Notice

Copyright (C) WIDE Project (2004). All Rights Reserved.