iSCSI 型ネットブートシステム起動性能評価システム

Development of a performance evaluation system for iSCSI-based thin client PCs system

本田　修啓

nhonda@gw.fukushima-u.ac.jp

福島大学総合情報処理センター

Fukushima University Information Network Center

概要

ネットブートシステムは,大学における実習教室の端末として多く採用されている.数百台の端末へのソフトウエアアップデートおよびセキュリティ更新を短時間で効率よく実施できる一方,起動時間に長い時間が必要となる問題が指摘されている.導入時の試験で実運用に近い環境でのテストは難しく,これらは運用後に表面化あるいは深刻化することが多い.本学総合情報処理センターでは,平成 23 年度導入した,iSCSI ネットブート型シンクライアントシステムの起動性能を MRTG グラフにより可視化し見やすく表示するシステムを構築し障害状況把握およびその解決に役立てているので紹介する.

キーワード

シンクライアント,ネットブート,可視化,性能評価,実運用

1. はじめに

福島大学総合情報処理センターでは,パーソナルコンピュータ端末約 300 台の Windows7 ネットブートシンクライアント端末（以下ネットブート端末）を平成 23 年 2 月に導入し運用している.セキュリティ維持や多数のクライアント端末の効率管理が得られる一方,以下のような障害報告がユーザから苦情として指摘された.

(1) 起動待ち時間増加
(2) 利用中の予期せぬ動作停止

アプリケーション利用中の予期せぬ動作停止は,Blue Screen Of Death (BSoD) と呼ばれる現象で,突然画面が青色になり,再起動以外に回復する手段はない.ファイル編集中に発生すると,利用者の被害は大きい.起動待ち時間の増加は,講義での実習開始時の影響が大きい.

特定の個人のみログオン後利用可能にならない状況は,ユーザプロファイル破損が原因であることが多く,個別にプロファイル初期化対応が必要で,管理者負担の増加の要因ともなっていた.

これらの不具合は突発的あるいは負荷集中時に発生し,導入時性能試験やユーザが少ない時間帯では障害が再現せず,被害を受けた利用者が障害報告を行わない場合も多く,障害の全体的な規模,起動遅延時間や発生時刻等ログやイベント確認に必要となるデータを把握することが難しい.このことが,迅速で適切な障害対策が行えない原因の一つとなっている.

報告するシステムは,ネットブートサーバ情報,クライアント接続スイッチ情報,iSCSI ストレージ情報を SNMP 等で収集し,日常の運用における,ネットブート端末の状況を収集し,管理者が状況を把握しやすいように可視化し,関連情報を並べて表示するシステムである.以下の情報をブラウザで簡便に確認可能としている.

(1) OS 起動後ログオン可能となるまでの時間
(2) クライアントと収容スイッチのリンク情報
(3) 起動されているクライアント数
(4) ネットワークトラヒック流量
(5) iSCSI 接続セッション数
(6) NAS 接続セッション数

2. 導入 iSCSI ネットブートシステム

評価対象システムは,NTT 情報流通プラットホーム研究所開発 STRAGEX を利用した iSCSI 型ネットブートシステムである.このシステムの起動時の動作を図 1 に示す.

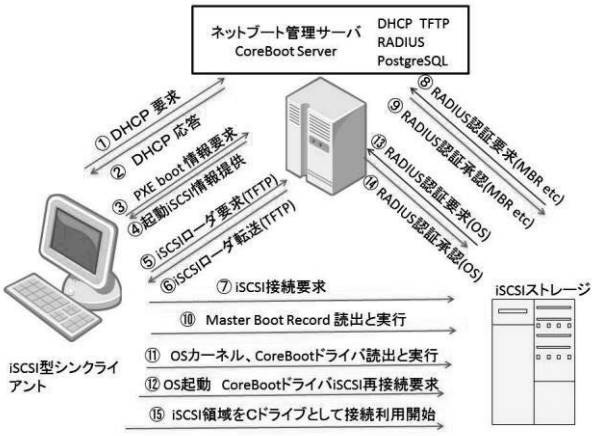

図 1 STRAGEX ネットブート起動時の動作

ネットブート管理サーバはクライアント起動時に起動のための情報を提供するのみで,クライアントは,iSCSI ストレージと直接通信し,OS およびアプリケーションを起動する.

起動後ネットブート管理サーバは,クライアントのログサーバとしてのみ動作し,負荷が原理的に少ないことから,クライアントシステム領域の配信サーバも兼ねるタイプのネットブート管理サーバと比して,より多数のクライアント管理を行うことができる特徴を持つ.

3 者のネットワーク構成を図 2 に示す.クライアントとネットブート管理サーバ,iSCSI ストレージは,同一 IP ネットワーク配置とし,2 個のレイヤスイッチを経由する.ジャンボフレームを使用しており,クライアントは⑪の OS 起動以降ジャンボフレーム(MTU 9000)で iSCSI ストレージと通信を行う.

ただしブート時は,0.5K あるいは 4K のブロック型転送が多く,ジャンボフレームの効果は限定的である.

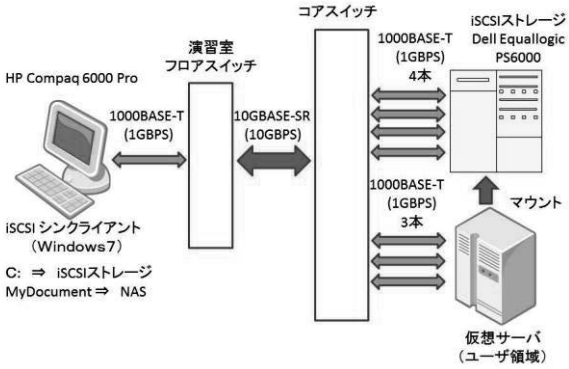

図 2 ネットワーク構成

2.1. 利用開始時トラヒックと起動時間

非混雑時クライアント起動時のトラヒックを図 3 に示す.電源ボタンを押し,ログオン可能になるまでの時間が約 110 秒あるが,うち 35 秒は OS 起動前の PXE 等の処理である.ログオン後利用可能状態まで約 1 分必要とする.ネットブートサーバは,OS 起動開始時およびログイン可能状態のタイミングを,内蔵データベースに記録する.起動時に流れるデータ量は約 1GB 弱である.

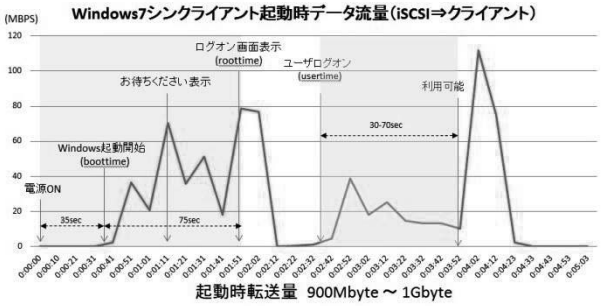

図 3 クライアント起動時データ流量(非混雑時)

文献[1]記載のように,ネットブート管理サーバは,起動時の処理のタイミングを内部データベースに保持する.この情報を利用することで,起動時刻,ログオン可能時刻,起動に要した時間をクライアント起動毎に把握することが可能である.

2.2. 混雑時トラヒックとクライアント起動

ネットブート端末 100 台同時利用時の演習室のフロアスイッチとコアスイッチ間トラヒックを図4に示す.帯域

は 10GBPS で余裕があるが,クライアント起動に要する時間が 5 分を超える遅延が多く発生する.

この状況は 100 人演習室 2 室で 200 人を超える利用者が一斉に利用開始するタイミングで発生する.

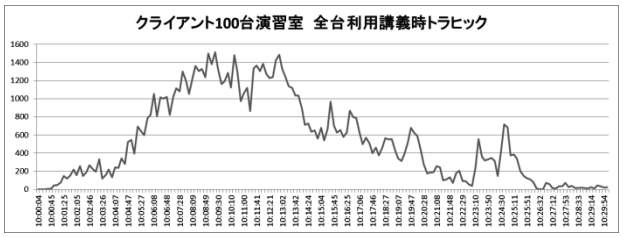

図 4 混雑時の演習室データ流量

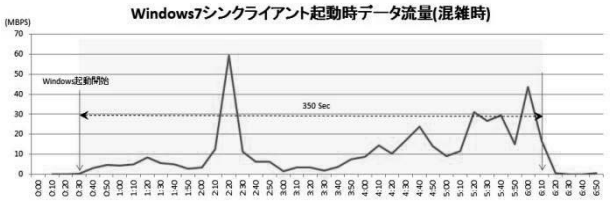

図 5 クライアント起動時データ流量(混雑時)

iSCSI ストレージ専用監視ソフトウェアで得られる情報を図 6 に示す.IOPS (Input/Output Per Second)が上限に達したことが確認できる.

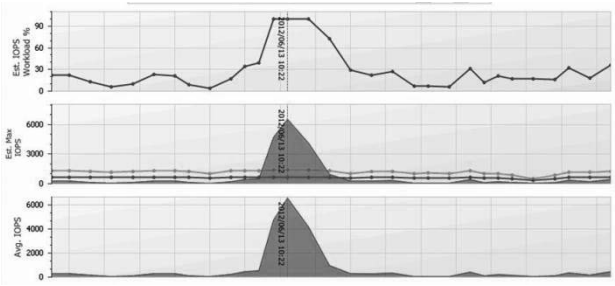

図 6 iSCSI ストレージ IOPS 状況

2.3. 100M リンク時のクライアント起動

BSoD の原因を調査する中で,クライアント起動時,フロアスイッチとの接続はオートネゴシエーションで 1GBPS 接続されるべきところ,まれに 100MBPS 接続になることが確認され,BSoD の原因の可能性を検討し,データ取得することにした. 非混雑時 100M 接続であると起動が遅くなる(図 7)ことでも検出は可能だが,フロアスイッチの SNMP 情報を確認することで混雑時でも確実に発生状況を把握することができる.(表 1)

表 1 リンク速度異常端末一覧

端末名	演習室名	boottime	roottime	usertime	stoptime	iscsi	linkspeed	起動時間
mul8022	MUL	2012-04-02 14:22:04	2012-04-02 14:25:04	2012-04-02 14:26:53	2012-04-02 15:12:17	str2_eth0	100	00:03:00
mul8016	MUL	2012-04-04 18:21:26	2012-04-04 18:24:35	2012-04-04 18:26:38	2012-04-04 19:01:03	str2_eth3	100	00:03:09
mul8016	MUL	2012-04-04 19:01:30	2012-04-04 19:04:22	2012-04-04 19:06:25	2012-04-04 19:07:29	str2_eth3	100	00:02:52
mul8013	MUL	2012-04-05 10:46:38	2012-04-05 10:49:45	2012-04-05 10:51:38	2012-04-05 11:49:11	str1_eth1	100	00:03:07
mul8016	MUL	2012-04-05 10:46:45	2012-04-05 10:49:56	2012-04-05 10:52:13	2012-04-05 11:09:10	str2_eth2	100	00:03:11

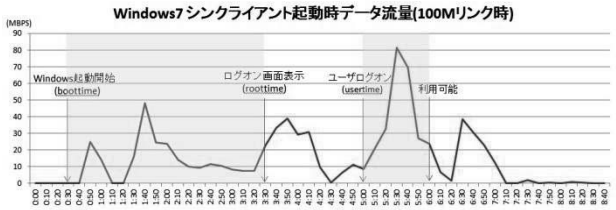

図 7 クライアント起動時データ流量(100M 接続)

3. 構築システム

3.1. 構成

構築した評価システムの構成を図 8 に示す.

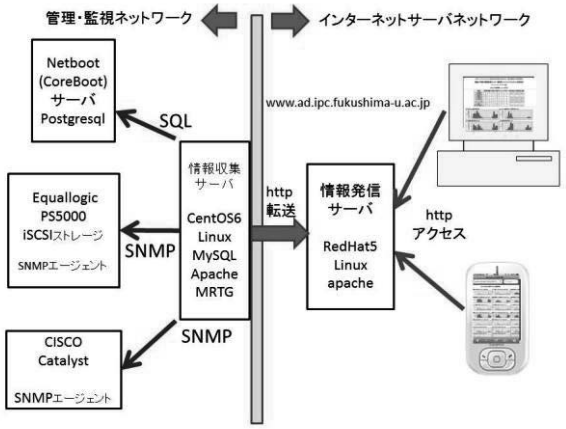

図 8 評価システムの構成

情報収集サーバは表2に示すように,ネットブートサーバ,iSCSI ストレージおよびフロアスイッチから,SQL,SNMP でデータ流量を除く,クライアント起動情報,クライアント接続情報,クライアントネットワークリンク情報を定期的に取得し,それを情報収集サーバデータベースに登録している.

表 2 収集情報一覧

収集情報	情報源	情報提供機器	プロトコル
OS起動開始時刻	iSCSIセッション情報MIB	iSCSIストレージ	SNMP
	ネットブート管理サーバ内部DB	ネットブート管理サーバ	SQL
利用可能時刻	ネットブート管理サーバ内部DB	ネットブート管理サーバ	SQL
接続リンク情報	演習室フロアスイッチMIB	演習室フロアスイッチ	SNMP
利用者ログオン時刻	ネットブート管理サーバ内部DB	ネットブート管理サーバ	SQL
利用終了時刻	ネットブート管理サーバ内部DB	ネットブート管理サーバ	SQL
	iSCSIセッション情報MIB	iSCSIストレージ	SNMP
iSCSIデータ流量	iSCSIセッション情報MIB	iSCSIストレージ	SNMP
演習室セータ流量	コアスイッチMIB	コアスイッチ	SNMP

収集データは,データベースから SQL,ネットワーク機器および iSCSI ストレージから SNMP で読み出し,MRTG でグラフを作成する.これらの処理に加え,関連するグラフ情報を並べて表示する html を定期的に自動作成するプログラム,情報発信サーバに転送するプログラムでシステムを構成している.プログラムは Perl で記述した.

情報発信サーバはインターネットに情報を発信する既存サーバ(www.ad.ipc.fukushima-u.ac.jp)を利用しており,アクセス制限は行っていない.学外運用支援スタッフの便宜,および学内スタッフが監視端末を離れても,スマートフォン等から閲覧できる便宜を優先させた.(図9)なお,発信情報はシステム性能情報のみであり,利用者アカウント名等,個人情報類は一切含ませていない.

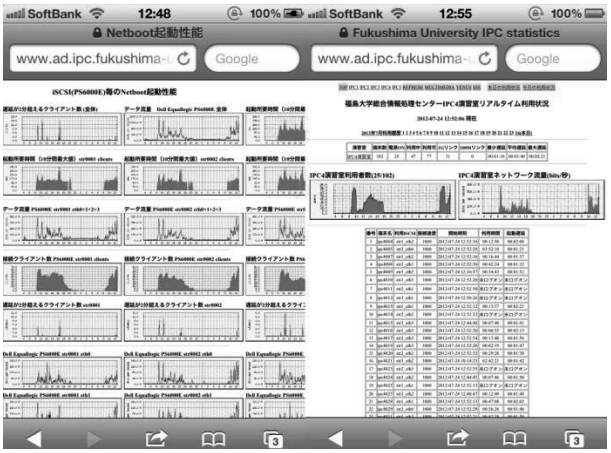

図 9 スマートフォンによる閲覧

3.2. 表示データと効果

3.2.1. 起動遅延情報

シンクライアント起動に要する時間情報は,ネットブートサーバ内部データベース情報の「OS 起動開始時刻」および「ログオン画面表示時刻」から計算して求めている.OS 起動前の時間は含まれない.したがって電源 ON 後ログオン画面表示されるのに要する時間は,この値に 30 秒程度加えた値になる.目視による実測およびパケットキャプチャログ確認を行った結果,ほぼ正確な値が得られることを確認した.

日常の利用で,どの時間帯に遅延がどの程度集中するか,どの iSCSI ストレージに集中するか確認が容易であることから,複数 iSCSI ストレージに起動クライアント割り当てを効果的に随時修正することが可能となっている.また修正の効果も確認しやすい.

3.2.2. 端末毎の起動に要した時間情報

起動し利用されているシンクライアント一覧を表示し,起動に要した時間情報も併せて表示する.加えて,最大値,最小値,平均値も表示する.

番号	端末名	利用ISCSI	接続速度	開始時刻	利用時間	起動遅延
1	ref6001	str1_eth3	1000	2012-07-06 16:22:09	00:19:57	00:01:35
2	ref6003	str1_eth0	1000	2012-07-06 16:12:05	00:57:06	00:01:28
3	ref6006	str1_eth2	1000	2012-07-06 16:36:07	00:08:23	00:01:32
4	ref6008	str1_eth0	1000	2012-07-06 16:34:04	00:10:01	00:01:27
5	ref6009	str3_eth0	1000	2012-07-06 14:12:26	08:40:13	00:01:06

演習室	端末数	電源ON	利用中	利用可	1Gリンク	100Mリンク	最少遅延	平均遅延	最大遅延
リフレッシュ	9	5	5	4	5	0	00:01:06	00:01:26	00:01:35

図 10 起動に要した時刻表示

3.2.3. 日統計情報

日毎の起動に要した時間情報をとりまとめる機能である.過去の日の情報も閲覧することができる.端末毎,利用毎の情報の他,最大値,最小値,平均値も表示する.

番号	端末名	利用ISCSI	接続速度	開始時刻	終了時刻	利用時間	起動遅延	備考
1	ref6001	str1_eth2	1000	2012-06-25 08:21:09	2012-06-25 08:29:58	00:06:30	00:01:22	
2	ref6001	str1_eth2	1000	2012-06-25 09:49:02	2012-06-25 10:00:42	00:09:21	00:01:25	
3	ref6001	str1_eth0	1000	2012-06-25 10:16:51	2012-06-25 10:40:02	00:20:36	00:01:38	
4	ref6001	str1_eth2	1000	2012-06-25 10:41:07	2012-06-25 11:38:32	00:55:05	00:01:28	
5	ref6001	str1_eth3	1000	2012-06-25 11:46:58	2012-06-25 11:56:53	00:05:40	00:01:25	
6	ref6001	str1_eth3	1000	2012-06-25 11:58:09	2012-06-25 12:17:03	00:16:04	00:01:35	
7	ref6001	str1_eth1	1000	2012-06-25 13:09:34	2012-06-25 14:36:40	01:24:33	00:01:35	
8	ref6001	str1_eth0	1000	2012-06-25 14:44:54	2012-06-25 15:43:38	00:55:58	00:01:50	
9	ref6001	str1_eth0	1000	2012-06-25 15:50:43	2012-06-25 16:07:39	00:14:32	00:01:27	
10	ref6001	str1_eth2	1000	2012-06-25 16:08:45	2012-06-25 17:04:45	00:53:07	00:01:59	
11	ref6001	str1_eth0	1000	2012-06-25 17:30:28	2012-06-25 17:42:38	00:09:40	00:01:25	
12	ref6002	str1_eth1	1000	2012-06-25 08:27:07	2012-06-25 08:42:10	00:12:31	00:01:26	

演習室名	利用回数	総利用時間	最小遅延	平均遅延	最大遅延	100MLink	異常終了	iSCSI-1	iSCSI-2	iSCSI-3
リフレッシュ	102	60時間49分	00:01:08	00:01:36	00:02:53	0	1	101	0	1

図 11 日統計情報

3.2.4. 月統計情報

演習室毎に,利用件数,月累積利用時間,起動に要した時間の最大値,最小値,平均値を表示する.

併せて,iSCSI ストレージ毎のクライアント接続数情報も表示させている.システムの利用統計データとしても活用可能な表示形式としている.

演習室名	利用回数	総利用時間	最小遅延	平均遅延	最大遅延	100MLink	異常終了	iSCSI-1	iSCSI-2	iSCSI-3
IPC1演習室	5599	6422時間52分	00:01:05	00:01:35	00:05:17	4	22	1722	1806	2071
IPC2演習室	725	796時間39分	00:01:14	00:01:56	00:05:47	0	5	0	0	725
IPC4演習室	2928	3939時間46分	00:01:06	00:01:44	00:04:31	0	14	923	987	1018
IPC5演習室	784	1017時間1分	00:01:06	00:01:35	00:04:00	0	5	302	465	17
IPCリフレッシュコーナー	1738	1206時間31分	00:01:04	00:01:35	00:04:23	0	5	1717	0	21
図書館マルチメディア室	1412	1895時間0分	00:01:05	00:01:28	00:03:36	0	14	625	782	5
IPC全体	13186	15280時間50分	00:01:04	00:01:37	00:05:47	4	65	5289	4040	3857

図 12 月統計情報

3.2.5. 接続速度情報

速度ネゴシエーションに失敗し 100MBPS 接続となったクライアント情報を記録し,各利用情報に合わせて表示させている.突発的な遅延原因の一つであり,クライアントNICあるいはネットワーク機器の交換により状況を改善している.ネットブートシステムのドライバに NIC リセットを追加することも若干の効果が見られた.

3.2.6. 起動端末数の可視化情報

MRTG を利用し,利用されているクライアント数の時間推移をグラフ化し表示させている.利用が多い時間帯,急激に利用が増加した時間帯を把握するのに有効である.

月,年グラフについてはピーク値も表示させ,平均化によってピーク値情報が失われないよう配慮している.

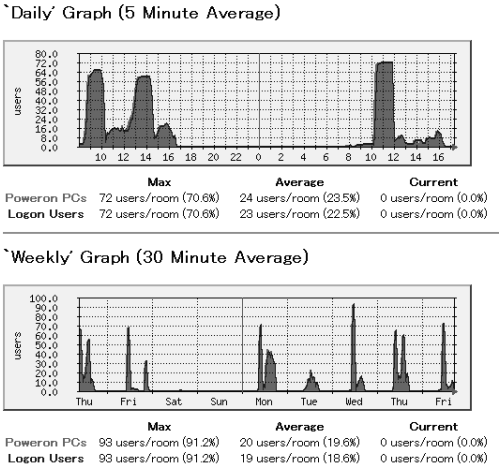

図 13　起動端末数情報

3.2.7. 起動に要した時間情報

ある5分間隔で,起動に最も長時間要したクライアントの起動時間を MRTG でグラフ可視化している.起動遅れが大きい時間帯の把握が容易である.

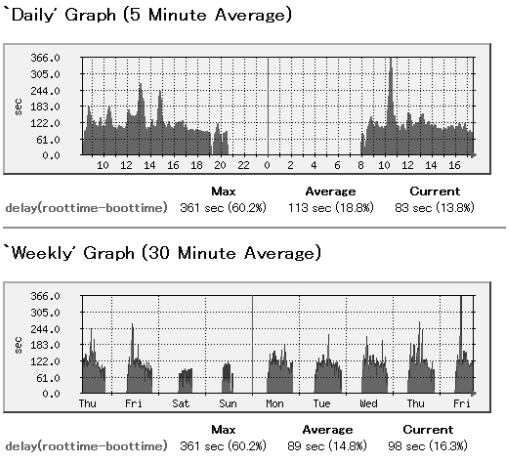

図 14　起動所要時間情報

3.2.8. 起動に長時間要した端末数情報

起動に 2 分および 3 分以上要したクライアント数を MRTG でグラフ化している.(5 分間のクライアント数)

起動遅れの発生時刻と規模が把握できる.

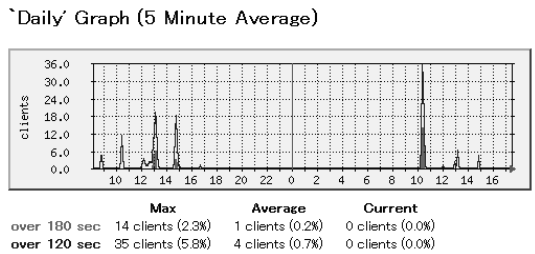

図 15　起動遅延クライアント数情報

3.3. ネットワーク流量情報

各演習室フロアスイッチとコアスイッチ間のネットワーク流量を MRTG 本来の機能でグラフ化している.

標準設定をそのまま利用しており,表示されるグラフは 5 分平均値であり,10 秒毎のサンプリンググラフよりピーク時流量は少なく表示されるが,起動端末数グラフ,遅延端末数グラフと併せて表示することで状況把握に便利である.

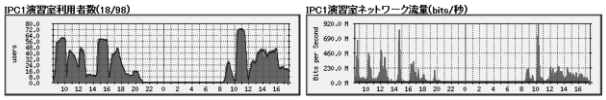

図 16　ネットワーク流量情報

3.4. iSCSI ストレージ情報

ストレージ状況は起動時間に大きな影響を与えると判断できるので,本システムでも SNMP を利用し,iSCSI ストレージインタフェース流量および,接続セッション情報から得られる起動端末数数情報,起動遅延情報等と同一画面に表示させている.

3.4.1. iSCSI 流量情報

iSCSI ストレージは3台あり,それぞれ4個の1000BASE-T インタフェースを有している.インタフェース毎の流量グラフの他,iSCSI 毎に各インタフェース流量を合算したグラフ,3 台の iSCSI の全流量を合算したグラフも作成させている. 図 17 は全合算流量である.

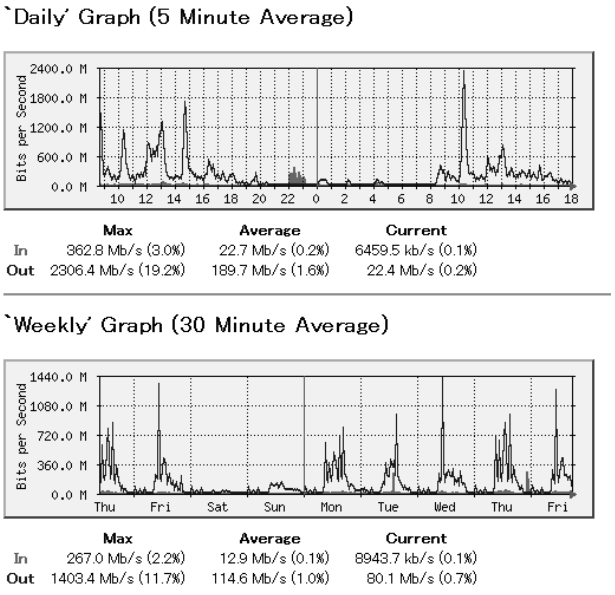

図 17　iSCSI データ流量情報

3.4.2. iSCSI 接続クライアント情報

iSCSI ストレージのクライアント接続 TCP セッション数を表として表示させている(図 18). クライアントが接続する iSCSI ストレージはネットブート管理システムで静的に指定するが,実運用の利用の偏り情報としての価値もあり,この情報を元にクライアントの iSCSI 割り当てを見直すことで,起動に要する時間を全体として低く抑えることが実現できた.

番号	端末名	利用ISCSI	接続速度
1	ipc1010	str1_eth2	1000
2	ipc1016	str1_eth2	1000
3	ipc1020	str2_eth3	1000
4	ipc1024	str3_eth3	1000
5	ipc1025	str1_eth3	1000
6	ipc1032	str2_eth1	1000

Equallogic	接続	Port	接続
STR0001	7	1	2
		2	1
		3	3
		4	1
STR0002	6	1	1
		2	2
		3	1
		4	2
STR0003	8	1	2
		2	1
		3	3
		4	2

図 18　iSCSI 接続クライアント数情報

3.4.3. iSCSI 接続クライアント数　可視化情報

iSCSI ストレージのクライアント接続 TCP セッション数を MRTG でグラフとして可視化した iSCSI ストレージ毎の接続クライアント数の変化と,同じ Web 画面に併せて表示される遅延可視化情報により,iSCSI クライアント数の変化とクライアント起動に要する時間の関係を直感的に把握することが可能となった(図 19).

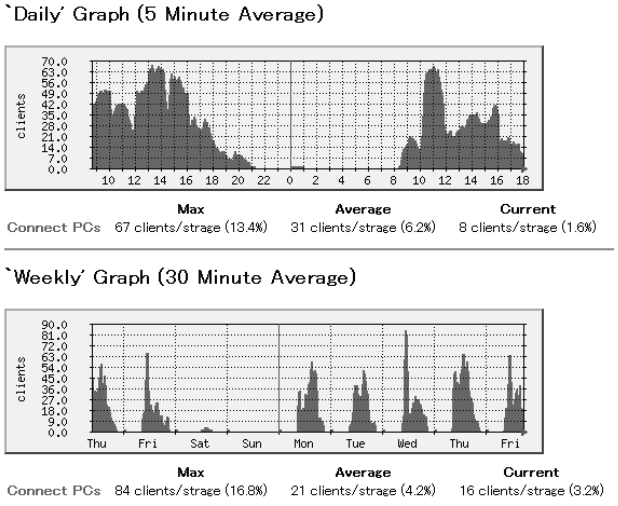

図 19　iSCSI 接続セッション数情報

4. まとめ

本システムは,導入当発生したプロファイル破損,BSoD の状況把握を目的に,順次機能を追加して行き,自動利用統計機能を加えて現在に至っている. 関係情報を MRTG, 発生した障害報告に対し,状況の把握を簡便に行うという目的は十分に果たしたと考える.関連する情報の同一画面表示は,状況把握効果が高く,自動統計機能は,遅延が大きい端末台数とその発生時刻,発生場所,発生規模(遅延が大きい台数) を記録し,行った障害解決の有効生を確認するのに有効であった.

プロファイル破損,BSoD については,iSCSI ストレージの予期せぬセッション断,ウィルス対策ソフトの予期せぬ動作,NAS 資源不足等が原因とわかり,改善され,障害報告および苦情は 0 に近い状態となっている.

今後は,負荷集中時の起動遅延問題の原因究明と解決に取り組みたい.

参考文献

[1] 市川俊一,岡純一,鷺坂光一：iSCSI を利用したシンクライアント PC システム STRAGEX,情報処理学会論文誌 : コンピュータシステム,Vol47 No.SIG12(ACS 15)

[2] 城所弘泰,村上登志男,磯上貞夫,坂本幸治郎 : シンクライアントの性能評価と大規模システムへの導入可能性の研究,学習院大学計算機センター紀要,Vol.29 2009

[3] 株式会社 NTT データ CoreBoot パンフレット http://www.coreboot.jp/wp-content/themes/coreboot/img/common/pamphlet.pdf

[4] 本田修啓：仮想サーバとクラウドサービスを活用した演習室クライアントシステム構築の一例,学術情報処理研究,No15 2011