BANCS 接続システム（4）
――大規模トランザクション処理と高可用性を実現した BANCS センタ接続

Interconnection between BANCS system and BANCS Center on ES 7000/W2KDCS

綾 野 昌 史

要 約 今回の BANCS システムは，性能要件，信頼性要件，可用性要件そして保守性要件と高品質な条件が要求されている．本稿では，BANCS システムと BANCS センタとの回線接続におけるインタフェース部分に着目し，特に

1．可用性要件の解決策として BANCS サーバをはじめ通信機器全てに完全二重化の相互バックアップ形態を採用

2．性能要件における弱点カバーの対策

の二つに焦点を当て，ミッションクリティカルなシステムである BANCS システムをどのように実現したかを事例として紹介する．

Abstract The BANCS system recently implemented requires the high level of processing performance, reliability, availability and maintainability. This case study report discusses how the mission critical BANCS system was implemented, paying our attention to the interconnecting interface between the BANCS system and the BANCS center, especially focusing on :

1. Adoption of full-duplex mutual backup configuration for BANCS servers and communication devices in order to satisfy the availability requirements.
2. Counter-measure taken to cover any weakness in performance requirements.

1. はじめに

本稿では，BANCS システムと BANCS センタとの回線接続におけるインタフェース部分に着目し，Unisys Enterprise Server ES 7000（以下，ES 7000）と MS 社の Windows 2000 Datacenter Server（以下，W 2 KDCS）上での回線接続におけるハイボリュームトランザクション処理と高可用性を実現したミッションクリティカルシステムに供するための方策と評価結果について記述する．BANCS システムに関しては，本誌掲載の論文「BANCS 接続システム（1）プロジェクト成功の鍵：POC 実践報告（山岸重雄著）」（以下，山岸論文）に記述されている．

2. システム要件とその実現に向けた基本方針

本システムは，山岸論文の 3 章で記述されているように性能要件，信頼性要件，可用性要件そして保守性要件と高品質な条件が要求されている．ミッションクリティカルシステムである BANCS システム実現のため，BANCS センタとの回線接続において以下のような基本方針を策定した．

1） 可用性要件では，H/W（ハードウェア）を完全二重化のロードバランス構成とした．

ロードバランス構成のポイントは，BANCS サーバおよび通信機器全てを完全二重化の相互バックアップ（ACTIVE/ACTIVE）形態にすることにより，障害時全面ストップを回避することである．通信機器障害時の切換えは，BANCS サーバから監視と制御をすることにより自動的に行わせた．また，BANCS サーバと通信機器のロードバランス構成は相互に非同期で行われることも可能とした．

2）　性能要件では，以下の条件で 2003 年 3 月末日迄の想定件数を処理することを目標とした．

・処理件数：132 trx/秒

（事務量予測：39,531 件/10 分＝約 66 件/秒に対して上り下りを考慮）

・回線数：本番 8 回線/被災 4 回線×2 センタ（関東，関西）

※回線数は，事務量予測と BANCS センタから提示された回線能力に対する伝送効率 80% 比率計算にて BANCS センタ 2 拠点に各 8 回線接続する．

・伝送制御手順：B 改全二重手順方式

上記処理件数（132 trx/秒）は両センタ合計の値であるが，最悪のケースとしてはフェイルオーバ時にセンタ障害で被災運用（または被災運用中のフェイルオーバ）した場合，BANCS サーバ 1 台で 100 trx/秒を処理できるように設計した．これは通常 4 台のサーバは各々 33 trx/秒処理しているが，フェイルオーバ時には 66 trx/秒を 1 台のサーバで処理する．この時被災運用に入った場合，1 センタの半分の処理量（33 trx/秒）が追加される．従ってサーバ 1 台で処理する処理能力は，99 trx/秒（公称：100 trx/秒）以上の処理件数が要求されるためである．

3.　開発システムの機能

3.1　ハードウェア構成およびその機能

BANCS センタとの回線接続には，通信制御装置である UST（Universal　Station Terminal）が必要である．今回，耐障害性向上のため BANCS サーバと同様に UST も複数台構成となり，その相互接続において以下のような検討を行った．

3.1.1　設計上の検討内容

当初，本番用 UST を 2 台，予備用 UST を 1 台で設計していたが，LAN 接続用 SW-HUB（スイッチング・ハブ）の接続方式で冗長構成上次のような問題が発生した．まず 2 台の SW-HUB を二重化して接続したところ，SW-HUB 障害時本番 UST が 2 台共使用不可能状態になる（図 1 参照）．また，SW-HUB を 3 台にすると BANCS サーバに 3 枚の LAN ボードが必要となる．この状況を完全に解決するためには，SW-

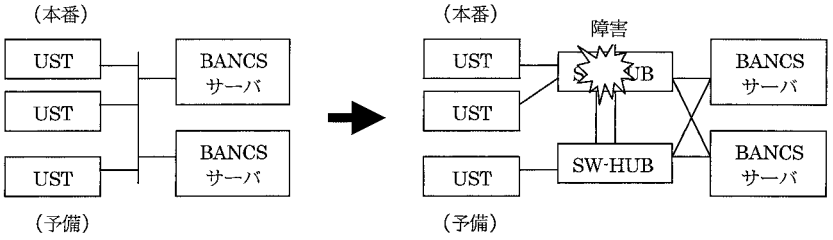


図 1　LAN 接続展開図（パターン 1）

HUB を 5 台で 2 階層に接続せざるを得なくなる（図 2）．しかし，この構成は機器の冗長度，障害の複雑性，コスト負荷の観点から最適とはいえなかった．そこで，UST を 2 台にし，相互に予備回線を持たせたロードバランス構成に方針変更をした．この方法で SW-HUB との接続関係を簡素化することに成功した（図 3）．

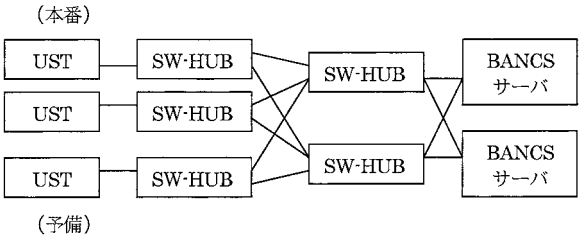


図 2　LAN 接続展開図（パターン 2）

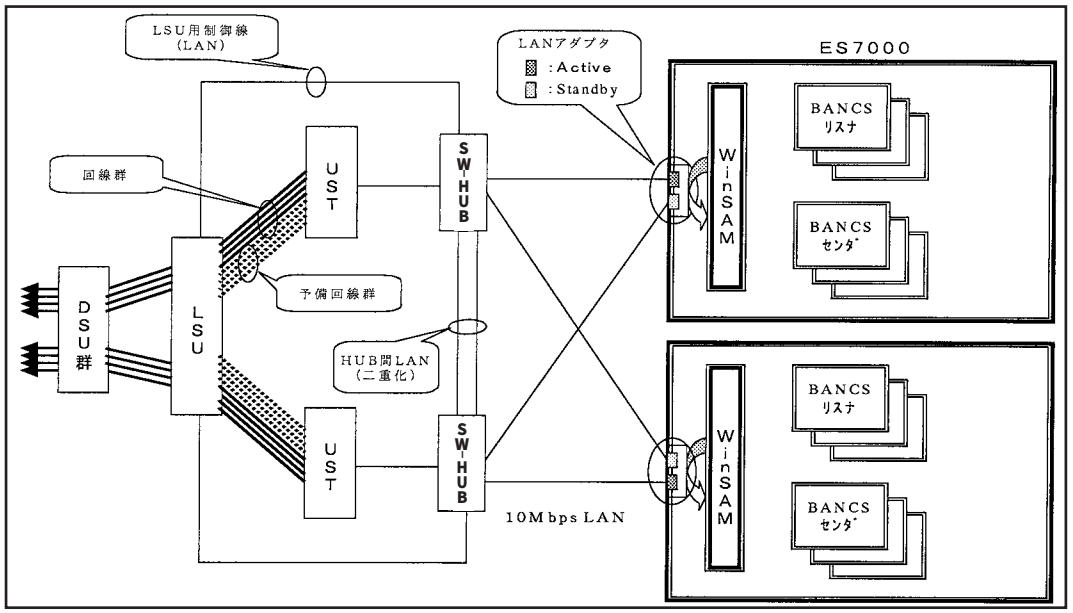


図 3　LSU-UST-サーバ間の接続形態

3.1.2 構　成　説　明

BANCS システムと BANCS センタは回線で接続され，LSU（回線切換え装置）を介し UST で収容した．UST と BANCS サーバ間は LAN 接続で TCP/IP にて受け渡しを行う．

以下では，冗長化構成実現のために採用した方策について述べる．

BANCS センタと BANCS サーバ間の接続は，H/W の冗長性を持たせるため，UST および SW-HUB を対の構成とした．また，SW-HUB 間の LAN ケーブルも冗長化のため二重化した．さらに BANCS サーバ側の LAN アダプタについても，2 枚の LAN ボードを一つの IP アドレスに割り当て，片方のみを使用する ACTIVE/STANDBY 構成とした．

SW-HUB 障害時には自動的に STANDBY 側の LAN ボードに切換える設計とし，また，UST 障害時には LSU（回線切換え装置）を使用し，もう一方の UST 内予備回線に切換える相互バックアップ形態を採用した．

UST 内のインタフェースボード（以下，I/F ボード）に収容する回線（最大 4 回線）は，本番回線と予備回線を混在分散させ，I/F ボード障害時の影響を最小限に抑えた．

各 UST は，各 BANCS サーバから同時に接続可能な状態にしておき，サーバ障害時には両 UST とも正常な BANCS サーバへ動的に接続を変更する機能を提供した．

また，UST 障害と BANCS サーバ障害は，どちらも非同期に起こりえることを考慮した障害対策を実現した．

3.2 UST（Universal Station Terminal）

UST は，TCP/IP プロトコルから B 改全二重通信方式への変換またはその逆を行うことのできるセイコー・プレシジョン社製のプロトコル・コンバータ（SC-8270）を採用した．しかし，標準の UST ではシステム要件を満足することはできず，個別改造を加える必要があった．1 点は半二重から全二重プロトコルへの対応による通信回線使用率向上の改造で，もう 1 点は 2 台の BANCS サーバに対する柔軟な制御機能の追加である．特に後者については，元来 UST はサーバ 1 台としか通信できないため，内部構造の改造から始めた．また，BANCS サーバの片方が障害を起こしても他方の BANCS サーバとは通信を継続させる必要があるため，通信先の BANCS センタに障害を検知させないことも重要である．以下に主な改造項目をあげる．

① B 改半二重から B 改全二重プロトコルへの対応．

② 2 台の BANCS サーバに電文を均等に振分けるため，電文内先頭 2 バイトの経路番号から振分先の BANCS サーバを算出する機能．

③ BANCS サーバ障害時に電文の振分先をフェイルオーバ先の BANCS サーバに切換えるため振分先 IP アドレスを正常なサーバから変更できる機能．

④ 入力電文が BANCS サーバに送信不可能な時でも，BANCS センタへは即時に肯定応答（伝送制御レベル）する機能．但し，入力電文は UST 内で破棄する．もともと UST は BANCS サーバへの送信が完了するまで無応答（伝送制御レベル）となるため，BANCS センタ側が回線障害を認識し回線閉塞することを回避する必要がある．

⑤ 上記同様，BANCS サーバに送信不可能な時でも，回線クローズさせない機能．他方の BANCS サーバは該当回線で取引を継続する必要がある．

⑥ プロトコル上，「TEXT 送信後無応答，応答督促，規定回数オーバ」時は，異常完了でなく，不能完了に変更する機能．これは，異常完了時のみ別回線へ迂回し，不能完了時は電文を破棄して電文の二重送信を避ける（欠送方式）ためである．

⑦ ラッシュ時入力電文の滞留を避けるため，TCP タイマ値の範囲拡大．
（1 秒単位→100 ミリ秒単位）

3.3 LSU（Line Switching Unit）

LSU（回線切換え装置）は，当社製の LSU-Ⅱ Plus を採用した．LSU は制御部（LAN

アダプタを含む）と電源部が二重化されているため，LSU 自体の二重化は行わないことにした．ただし，LSU 内部には，複数の LSM（Line Switch Module）が搭載されているため，単独の LSM 障害時には全回線の半分しか影響を受けないように結線した．最悪 LSU 自体の障害（全回線障害）が発生した場合でも，LSU 両端のコネクタを取り外し，UST と DSU（回線接続装置）を直接結線することで解決できる．

3.4 SW-HUB（Switching-HUB）

SW-HUB は，BayStack 社製の 450-24 T を採用した．特徴は MLT（Multi Link Trunk）機能が標準装備されていることである．これにより SW-HUB 間の接続に最大 4 本までの物理リンクを 1 本の論理リンクとして扱える．この間のケーブルを含め，断線検知には SNMP の通知機能を使用した．また，LAN 設定は，UST の仕様に併せ 10 Mbps で AutoNegociation を Disable にした．さらに LAN 上ループが発生しないことから SpanningTree は Disable にした．これは，LAN アダプタが自動的に切換える際の時間短縮となる．

3.5 LAN アダプタ

LAN アダプタは，Intel 社製のドライバを使用した．特徴は 2 枚の LAN ボードを Primary と Secondary に設定した Active/Standby 構成である．通常 Primary 側のみ使用され，断線等異常を検知すると自動的に Secondary に切換え，また復旧も自動的に行われる．

3.6 ソフトウェア構成とその機能

H/W の冗長構成と合わせ，耐障害性向上のため，ソフトウェアに関して次の機能を開発した．

① 振分先 IP 設定機能
② 回線オープン/クローズ機能
③ UST 稼働監視機能

3.6.1 振分先 IP 設定機能

UST との接続にはセイコー・プレシジョン社製の WinSAM（通信制御ソフト）を使用した．

今回 BANCS サーバが 2 台となるため，仮想 IP アドレスを設定し，サーバ障害時に IP アドレスを自動的に移動させることを検討した．この機能はクラスタサービスで提供されている．これに伴い，吸収サーバ側に WinSAM をさらに一つ起動することも検討した．しかし，1 台のサーバに複数の WinSAM を稼働させるようなマルチ対応はされていない上，一つの WinSAM で複数の IP アドレスを認識させることも困難なことが判明した．そこで，BANCS サーバには一つの WinSAM（一つの固定 IP アドレス）のみを稼働させ，各 BANCS サーバへの入力電文の振分けは UST 側で機能させるよう内部テーブルを持たせた．さらに BANCS サーバから動的に振分先 IP アドレスを変更できるようにした（3.2 節 UST 参照）．この制御を行うのが振分先 IP 設定プログラムである．

このプログラムが BANCS サーバ障害時に起動されると，全ての入力電文は他方の正常サーバに送信されるようになる．この際 UST に対してコマンドを送出するが，その応答は BANCS リスナに通知される．そこで，BANCS リスナとの情報交換用と

して UST 状況管理テーブル（共有メモリ）を利用した．

3.6.2　回線オープン/クローズ機能

WinSAM 経由で電文の送受信を行うのが BANCS リスナ/センダの機能である．

WinSAM はマルチスレッドの AP（アプリケーション）対応がなされていないため，回線数だけのプロセスを各々起動することにした．この場合，BANCS リスナのプロセス障害が発生すると，入力電文は WinSAM にて破棄される．これが頻発すると BANCS センタ側で AP レベルでの電文の応答タイムアウトが多発するため，回避する必要が出てきた．その対策が回線閉塞（回線オープン/クローズ機能）である．回線がクローズされると BANCS センタからの電文送信には無応答となる．これにより BANCS センタ側に該当回線を障害と認識させ迂回させることができた．一方回線クローズは，クラスタ側の監視プログラムからリスナもしくはセンダ障害時（今回，3 回再起動しても障害になる時）に起動することを可能にした．ただし，BANCS リスナもしくはセンダのプロセス障害が多発した場合，全回線がクローズする前にフェイルオーバさせた（通常運用時 4 プロセスでフェイルオーバ発動）．また，この時，吸収サーバ側のリスナもしくはセンダは正常稼働している場合があるため，フェイルオーバ時，有効回線（リスナおよびセンダ共に正常な回線）を再度オープンするよう制御した．ここで回線閉塞を使用せずリスナの 1 プロセス障害でフェイルオーバさせることも考えられるが，障害規模に対して業務の影響が大きいと判断し，上記のような決定となった．

3.6.3　UST 稼働監視機能

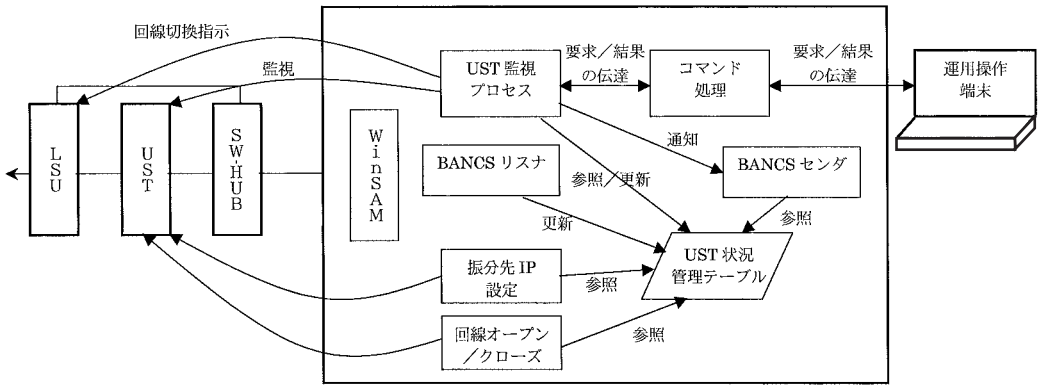
UST の障害を検知し，正常な UST に切換える機能が UST 監視プロセスである．

常時 UST に対して Ping を送信し，もし規定回数（今回，15 秒×4 回）以上の無応答を検知した場合は，UST 障害と判断し，LSU に切換えコマンドを発行する．この場合，BANCS センダの送信先 UST を切換え先 UST に変更させるため，情報交換用として UST 状況管理テーブルを使用した．LSU の切換えタイミングは，イベントを発行することにより各プロセスに通知した．さらに UST 監視プロセス自体の障害も想定し，監視対象は全 UST とした．UST からみると両 BANCS サーバから二重に監視を受けることになる．また，手動による LSU 切換えや UST の復旧等に各種コマンドも用意した．これら LSU への指示経路は 2 本の制御線の内 1 本を使用した．ここでもし一方の制御線経由で切換え操作に失敗した場合でも，他方の制御線で自動的に発行させるよう実装した．

UST 状況管理テーブルは，全 UST や LSU の構成情報を保持しているが，排他制御の煩わしさや整合性を維持する必要性からアクセス用独自 API を開発した． また，業務処理終了時の内容を翌日に引き継ぐために，全プロセスが UST 状況管理テーブルへのアクセス終了時，その内容をファイルに保存できるようにバックアップ機能も実装した．耐障害性向上の観点からのソフトウェア全体構成を図 4 に示す．

3.7　WinSAM の開発経緯

開発開始当初（2000 年 5 月），WinSAM は Windows 2000 版を新規開発中であり，バージョン V 5 M 0 のリリースを待って実際のテストや効率測定を行った．ところが，V 4 M 2 に比べ V 5 M 0 の処理能力が極端に悪く（V 4 M 2 に比べ半分程度），実用に



名　称	説　明	備　考
WinSAM	UST との間でﾃﾞｰﾀの送受信や制御を行う．	ｾｲｺｰ・ﾌﾟﾚｼﾞｼｮﾝ社製
BANCS リスナ	WinSAM 経由で UST からのデータを受信する．	
BANCS センダ	WinSAM 経由で UST へデータを送信する．	
回線オープン／クローズ	回線のオープンおよびクローズを実行する．	UST 電源 ON 時回線クローズ状態
振分先 IP 設定	UST に登録されているサーバの IP アドレスを動的に変更する．	BANCS サーバのフェイルオーバ時に起動
UST 状況管理テーブル	UST，LSU の状況を管理するテーブル	共有メモリ
UST 監視プロセス	UST を定期的に監視し，障害を検知した際に LSU へ切換えを指示する．	
コマンド処理	運用操作端末からのコマンドを解釈し，実行結果を返す．	

図 4　ソフトウェアの全体構成

絶えられないと判断した．そこでセイコー・プレシジョン社と協議の結果，V4M2 の内部構造をベースに見直しを実施し，シングルプロセス対応ながらマルチスレッドで稼働する構造に変更したものを V5M1 として再構築した．幸いこのバージョンで効率測定を行った結果，要求を満足することができた．また，その後も機能向上や安定性向上としてバージョンアップ要求を行い，現在 V5M1SE（PTF01）を客先リリースし，本番を迎えた．WinSAM のリリース時期とバージョンを表 1 に示す．

表 1　WinSAM のリリース時期とバージョン

リリース時期	バージョン	説明
2000 年 5 月時点	V4M2	Windows 95,98 対応(ｵﾘｼﾞﾅﾙは UNIX 対応版)
2000 年 10 月	V5M0	Windows 2000 Server 対応(ｵﾌﾞｼﾞｪｸﾄ指向対応版)
2001 年 4 月	V5M1	Windows 2000 Server 対応(V4M2 ﾍﾞｰｽ改訂版)
2001 年 6 月	V5M1 Second Edition	同上（効率関連ﾊﾟﾗﾒﾀ追加版）
2001 年 8 月	V5M1 Second Edition (PTF01)	同上（ターミナルサービス対応版）

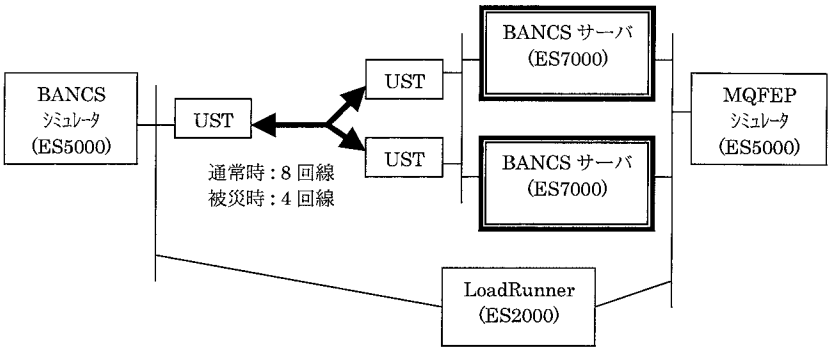
4.　評価・検証について

効率測定の環境としては，BANCS センタ想定機を BANCS シミュレータに，勘定系システム想定機を MQFEP シミュレータに見立て，双方から電文の送受信を行わせた．電文をラッシュさせるために，高負荷ツール（LoadRunner）を使用した（図 5）．

4.1　前　提　条　件

①　通常時とフェイルオーバ時：通常 8 回線

② 被災時と被災＆フェイルオーバ時：通常 8 回線＋被災 4 回線＝12 回線
③ 被仕向電文と仕向電文比率：6 対 4
④ BANCS シミュレータ側の UST：応答割込み機能（4.2 章にて記述）ON
⑤ Affinity（プロセスを特定のプロセッサに関連付けることにより，プロセッサーキャッシュの再読込負荷を軽減させる方法）設定：ON



・BANCS サーバ：ES7000/W2KDCS×2 台
・BANCS および MQFEP シミュレータ・サーバ：ES5000/Windows 2000 Advanced Server
・効率負荷ツール・サーバ：ES2000/Windows 2000 Advanced Server + LoadRunner

図 5　効率測定環境

4.2 改 善 内 容

当初の測定結果は，システム要件で求められていた効率には程遠いものであったため，原因追及をしたところ，次の点が改善ポイントとなった．

1） WinSAM 自体の効率問題→WinSAM の改訂バージョン（V 5 M 1）にて対応

WinSAM は，Windows 2000 対応と同時にオブジェクト指向に基づく大幅な内部構造の変更を行ったが，効率が伴わず旧バージョンをベースとした構造に戻さざるを得なかった（3.7 節参照）．

2） 電文サイズによる TCP/IP の Nagle アルゴリズムによる影響
→WinSAM の環境設定変更

本システムの電文サイズは 100 バイト長であるが，セイコー・プレシジョン社の測定では 3000 バイト長を使用していたため，両者で効率測定結果に差異が生じた．そこで，測定環境や方法を調査したところ，TCP/IP は Nagle アルゴリズムを実装しており，小さなセグメントを蓄積して一つの大きなセグメントにすることが判明した．これは，送信するパケットの数を減らすことによりネットワーク使用率を低下させる効果がある．ただし，今回は Nagle アルゴリズムを無効にした方が効率向上したため，WinSAM の環境設定に Nagle 無効オプションを追加した．

3） TCP/IP の ACK 受信遅延→WinSAM の環境設定変更

UST は電文を送信した後 TCP の ACK 受信を待ってから次の電文を送信する．ところが，TCP プロトコルでは ACK が最大 100 ms～200 ms 遅れて出る場

合がある．これを防ぐために，WinSAM から ACK 相当の電文をダミーで送信することにより遅延回避を可能とした．

4） B 改全二重プロトコルの応答割り込み機能→UST の改造

応答割り込みは，伝送ブロック中に従局応答（ACK/NAK）を挿入することで伝送効率を向上させることができる．しかし，UST は応答割り込みを受信できても送信することができなかった．BANCS センタ自体には応答割り込み機能があるため，この機能を追加開発し BANCS センタ想定機側の UST に実装させた．一方，BANCS サーバ側の UST にも応答割込み機能を適用したかったが，BANCS センタの伝送制御試験に間に合わなかったため断念した．

4.3 結　　果

以上の改善をへて表 2 の結果が得られ，要求されたレベルをクリアできた．

表 2　測定結果

① 各サーバ状態の測定結果

		通常時	被災時	フェイルオーバ時	被災&フェイルオーバ時
要件定義	Trx/sec (SQL Server)	33.0	50.0	66.0	100.0
計測結果	Trx/sec (SQL Server)	35.3	50.9	68.9	100.8

② 被災状態での最高効率

		被災時
計測結果	Trx/sec (SQL Server)	110.6

なお，効率関連の詳細については，山岸論文を参照されたい．

5. お わ り に

今回重要視したのは，可用性要件と性能要件の 2 点である．この観点からすれば目的を達成できたと考える．一方メインフレームにおいては常識である通信制御装置の相互バックアップ構成機能であるが，オープン系においてこれらの構築は 1 から行わなければならなかった．また，効率におけるさまざまな問題対応や改善（WinSAM の再構築他）に約半年の歳月を要した．この点は開発時のリスクと見る必要があるだろう．

以上のように，本システムではメインフレームと同等の機能および性能を ES 7000 と W 2 KDCS の組み合せで実証することができた．このプロジェクトで培われたノウハウは，Windows ベースでのミッションクリティカルなシステムへの使用に充分耐えられるであろう．本稿が他のプロジェクト開発の参考になれば幸いである．

執筆者紹介　綾　野　昌　史（Masashi Ayano）
1958 年生．1980 年電気通信大学機械工学科卒業．1985 年日本ユニシス㈱入社．メインフレーム中心の客先システムおよび基盤開発を担当．1994 年以降官公庁向け入札案件の提案に従事．1999 年 Y 2 K 問題管理開発後 BANCS 開発に従事．現在，ファウンデーションサービス部ミドルウェアサービス室に所属．