DEOS-FY2009-PP-06

仮想化を利用したディペンダブルOSの構築

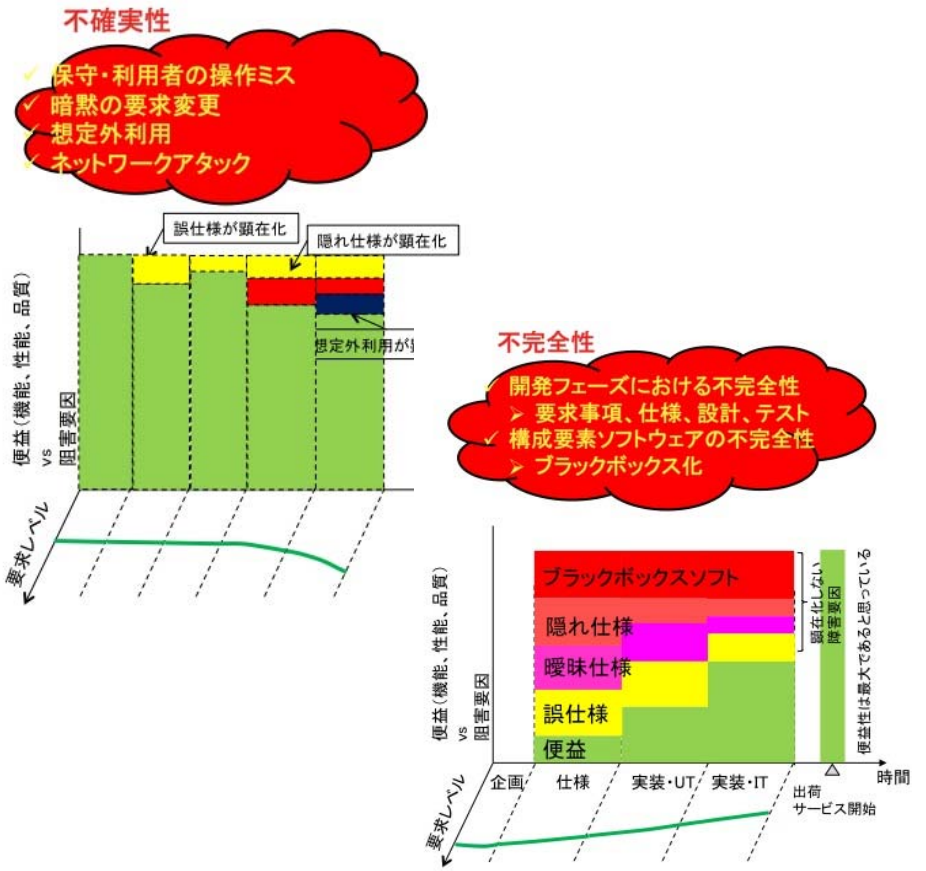
早稲田大学　基幹理工学部　情報理工学科
中島　達夫

目次

- 中島チームの研究概要
- 仮想化技術の有効性
- なぜ新しい仮想化技術が必要か？
- 高信頼仮想化技術の実現
- 研究の現状とデモストレーション
- 重要な研究成果
- 結論

中島チームが検討するオープンシステムディペンダビリティへの対応

- 仮想化技術を利用する　場合の不完全性・不確実性の影響の最小化
  - リスク管理手法を用いた仮想化技術
- 仮想化技術を利用することによる不完全性・不確実性の減少
  - OSの整合性回復機能
- 仮想化技術を利用することによるリソース管理の　柔軟性
  - 動的マルチコア仮想化



阻害要因	支援目的	フェーズ	対象機能
■自然現象 ■人的ミス ■悪意ある攻撃 ■Hardware故障 ■Software故障	■可用性 ■信頼性 ■安全性 ■整合性 ■保全性	■仕様 ■設計 ■実装・単体試験 ■総合試験 ■流通 ■運用 ■保守・更新 ■破棄・再利用	■ファイルシステム ■通信 ■実時間処理 ■電力・性能 ■システム全体

中島チームの研究・成果概要

システム安全構築技術

Mechanism Supporting Dependability (MSD)

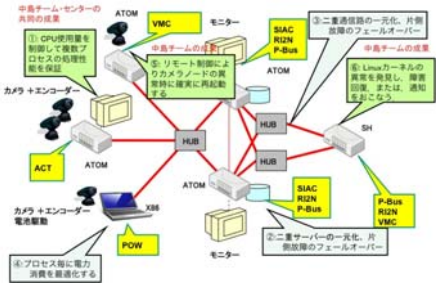
動作時間予約機構 (TR)
マルチコア電力制御機構 (PWR)
デバイス電源制御 (DEV)
耐故障ネットワーク機構 (RN)
チェックポイント・リスタート機構(CPR)
プロセスマイグレーション (PMIG)
シングルIPアドレス (SIAC)
ロギング &トレーシング
早期警戒型モニタ機構 (EWA)

MSD A (P-Component)
MSD A (P-Component)
• • •

Linux 基本拡張
•ディペンダビリティ機能拡張基盤：P-Bus
•ハード実時間処理＆マルチコア対応

Virtual Machine
OSディペンダビリティ支援
（OS異常検出、回復、セキュリティ監視・対策）

保守運用支援技術
- 耐ネットワーク故障
- 耐ノード故障
- 耐ソフトウェア故障
- セキュリティ攻撃防御
- 予兆検出

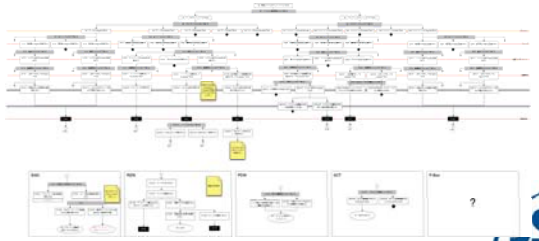


検証から漏れた
バグによる障害への対応

Linuxが想定していない
マルチコアプロセッサへの対応

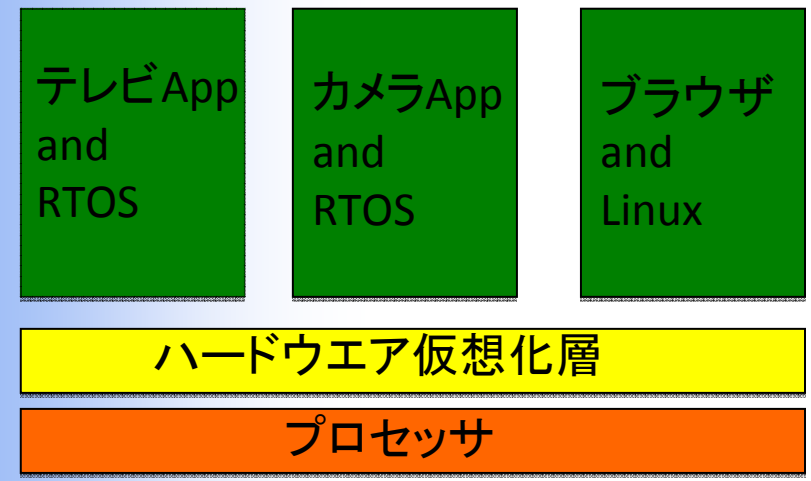
既存のRTOS資産の有効利用

DEOSコアチームにおける役割
1) 仮想化層をコンフィグレーションした場合の有効性の明確化
2) 仮想化技術を導入する場合のD-Caseの検討
3) 統合デモにおいて仮想化技術の有効性を実証

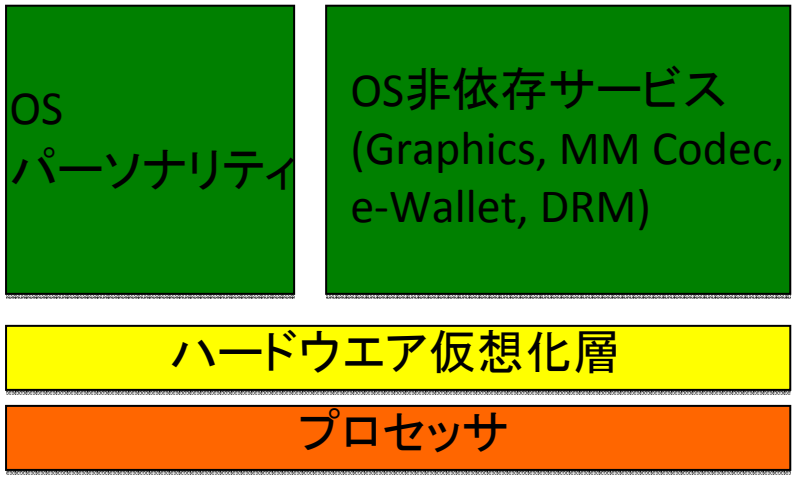


仮想化技術の有効性(複数OS)

- 複数のOSを同時実行可能とすることにより，様々な付加価値の提供が可能
  - 遠隔監視/メンテナンスによりOSを外部から再起動可能
  - 制御系のアプリケーションのRTOSを利用した高速再起動
  - 既存のRTOS上のアプリケーションのLinux上への移植コストの削減
- 仮想化技術を利用しないと，すべてのアプリケーションをLinux上で実行　する必要があるため，ディペンダビリティの保証が困難となることがある．



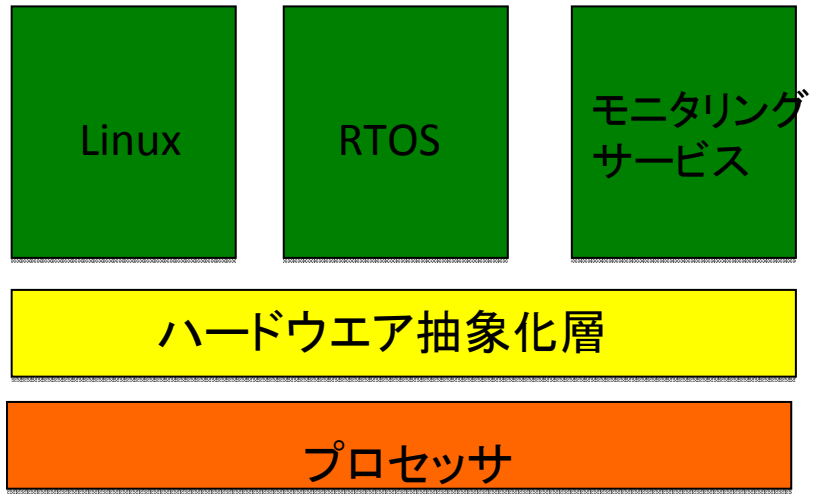
複数の機能をOSとアプリを組みで再利用可能とする



OS依存サービスを時代の流行と無関係にする

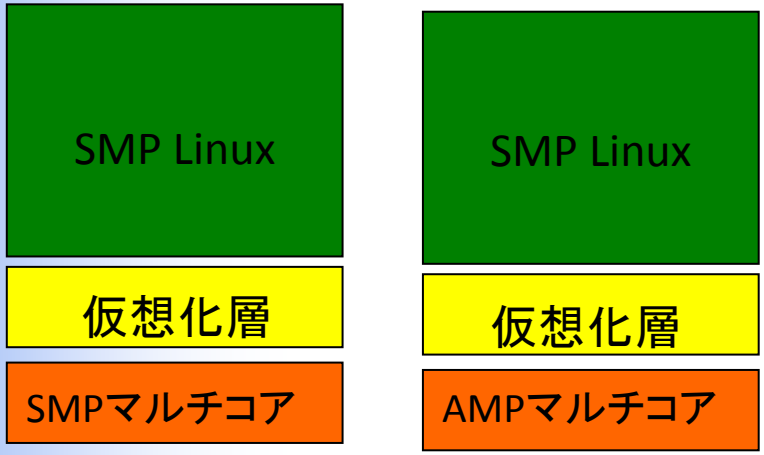
仮想化技術の有効性(モニタリング)

- モニタリング機能によるLinuxのディペンダビリティ向上
  - ミスコンフィグレーションによる間違ったカーネルモジュールの挿入を禁止
  - コードやデータ部の整合性の保証
  - システムコールテーブルの整合性の保証
  - カーネルがハングしていないことの検出
- 仮想化技術を利用しないと，Linuxの異常時でもモニタリング機能を実行出来るようにすることは不可能である.

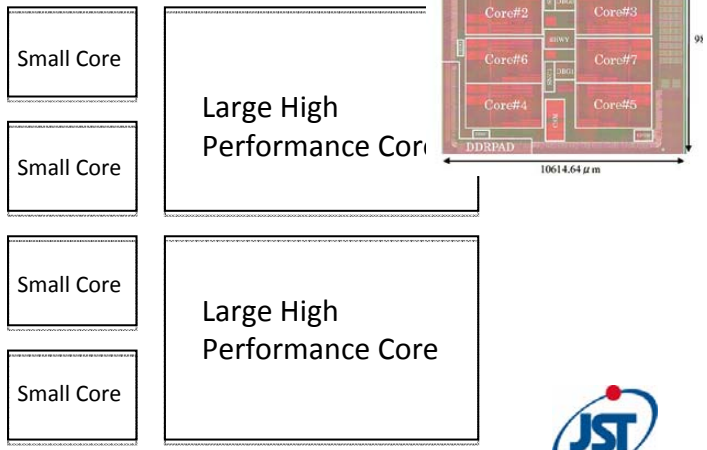


仮想化技術の有効性(マルチコア)

- リアルタイム性と消費電力の柔軟なトレードオフの実現
  - 各コアの電源のON/OFFをの仮想化層での支援により，　　アグレッシブな電力制御を可能
- マルチコアプロセッサの進歩に追従することを可能
  - 様々なヘテロジェニアスマルチコア
- 仮想化技術を使用しないと，将来のマルチコアプロセッサの進展に容易に追従することは不可能である．



AMPマルチコアの例



仮想化技術の有効性のまとめ

- 携帯電話，車載，デジタル家電等のLife Cycle Cost(LCC)の低減（複数OS, マルチコア）
  - 既存の資産を変更無しに有効に利用することが可能
  - マルチコアプロセッサの進歩に安価に対応することが可能
- 説明責任、メンテナンス性の向上（モニタリング）
  - 遠隔監視・制御により分散サービスの障害時の　対応が容易
  - モニタリングサービスの利用によるカーネル内の　整合性の崩壊からの回復
- 既存の仮想化技術は有効性1のみを支援

Cyber Physical Systems

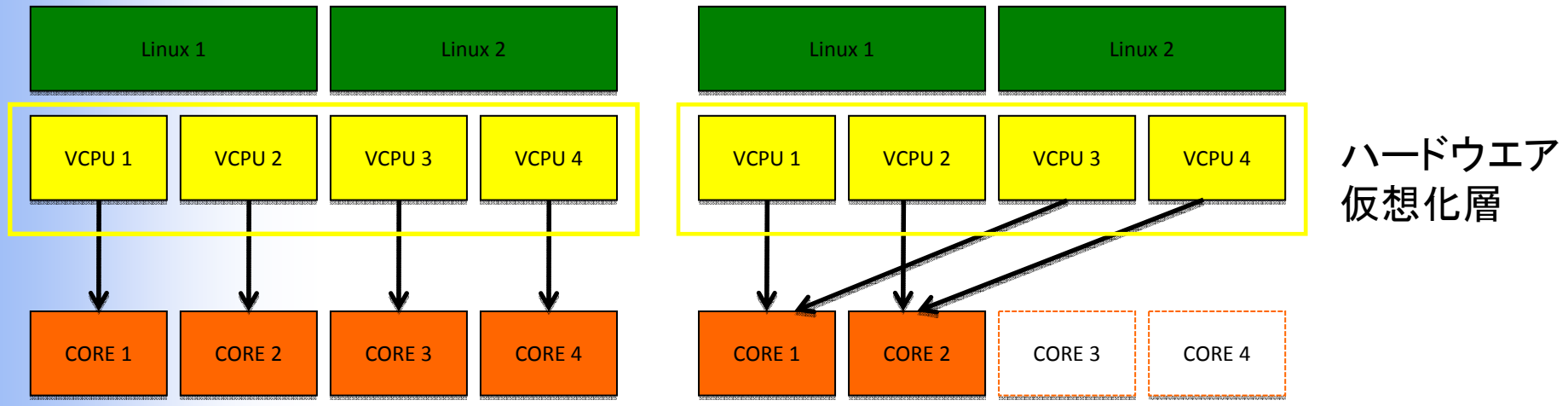
Air traffic control

Weapon control

Intelligent transportation

- 実行時の不完全性・不確実性の考慮
  - 仮想化技術を利用する時の不確実性の影響の低減
  - 仮想化技術を利用することによる不確実性の減少
- 柔軟なマルチコアプロセッサ対応
  - 実世界を状況認識するための安価な高性能処理の提供
  - ディペンダビリティのために他の非機能的要件を犠牲にしない
    .
    - リアルタイム性，性能，消費電力，セキュリティ間の柔軟なトレードオフ
- 開発する仮想化技術の特徴
  - 動的マルチコア仮想化アーキテクチャ
  - リスク管理手法を利用した仮想化層の実現
  - モニタリングサービスによるLinuxの整合性監視

動的マルチコア仮想化アーキテクチャ

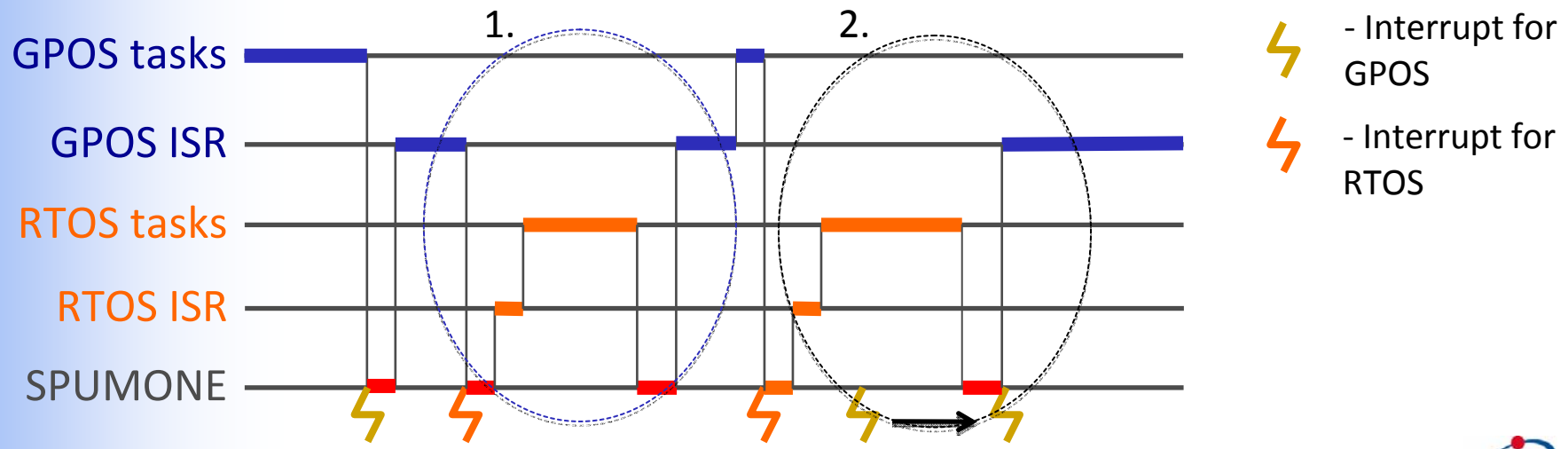


ゲストOSの変更の最小性

- 仮想CPUを実CPUより多くの数生成することを可能とすることで，柔軟な電力管理が可能
  - VCPUマイグレーションとスケジューリング
  - Linuxを変更せずに柔軟にコアのON/OFFが可能
  - リアルタイム性，性能，消費電力のトレードオフ

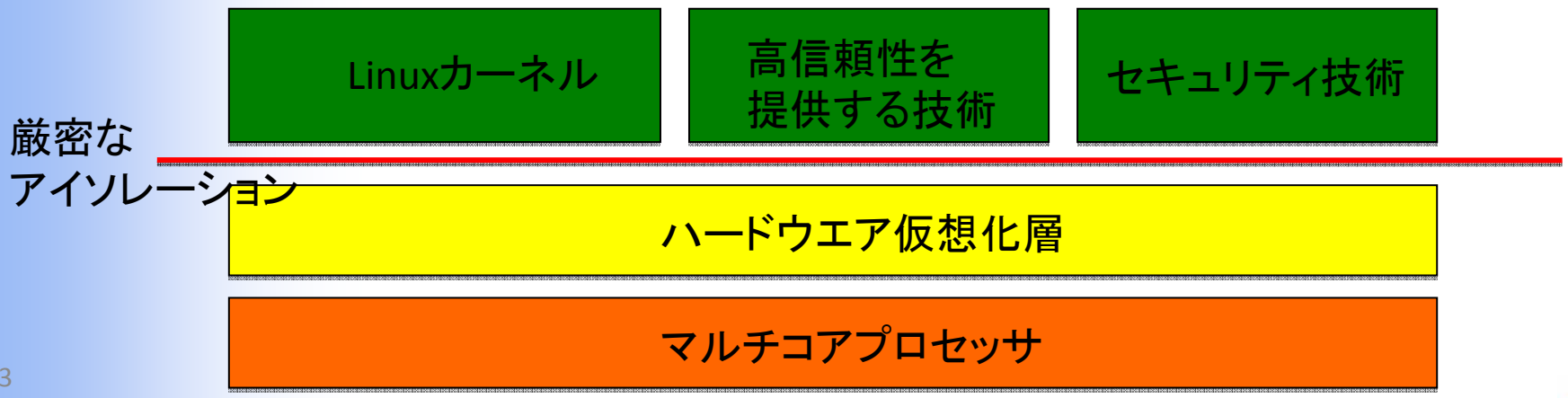
VCPUスケジューリングモデル

- 各OSのリアルタイム制約を保証するためRTOS上のアクティビティは汎用OS(GPOS)上の割り込み処理より高い優先度により実行する.
  - RTOSのタスクは汎用OSの割り込みハンドラを横取りする
  - RTOSのタスクは汎用OSの割り込みハンドラの実行を遅延させる
- 各コア毎に実行可能なVCPUのレディキューを持つ
  - VCPUマイグレーションはVCPUエンティティをコア間で移動する.

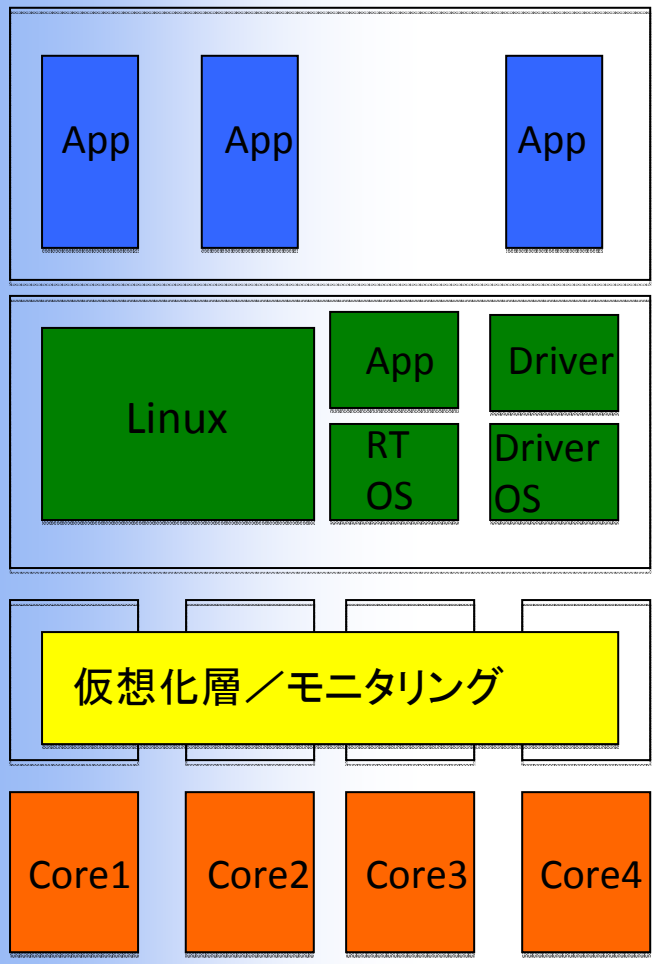


リスク管理手法を用いた仮想化層(VMC)の実現

- 既存のシステムでは，信頼性やセキュリティの実現にOSカーネルと独立した高信頼カーネル(Trusted Core)が一般的に利用
  - しかし、高信頼カーネルのバグがないことは仮定できない
- オープンディペンダビリティの実現のためには，理想的な高信頼カーネルを前提とするのではなく，様々な問題をリスクとして考え，それらへの対応を考慮することにより信頼性を向上する.



アーキテクチャとリスクの定義



ユーザスペース

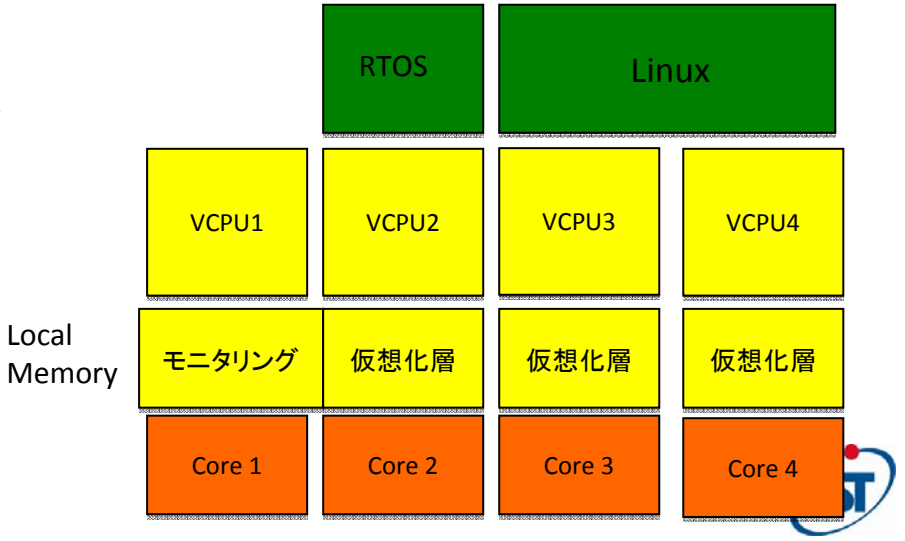
カーネルスペース

各コア毎のローカルメモリ

システム全体の簡単なFMEAライクな障害解析
- RTOSが壊れた場合
- Linuxが壊れた場合
-仮想化層が壊れた場合

多面的な再起動の利用
-障害の影響があるOSのみ独立に再起動
-重要な情報をチェックポイントとして保存することによる，再起動の高速化
-Linux内のエラーの仮想化

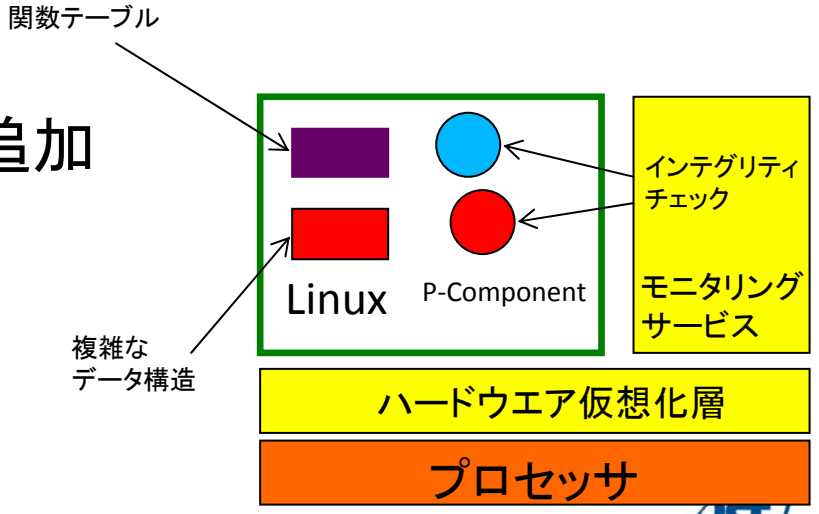
仮想化層と障害検出機構の信頼性の向上



モニタリングサービス(VMO)とPDCA

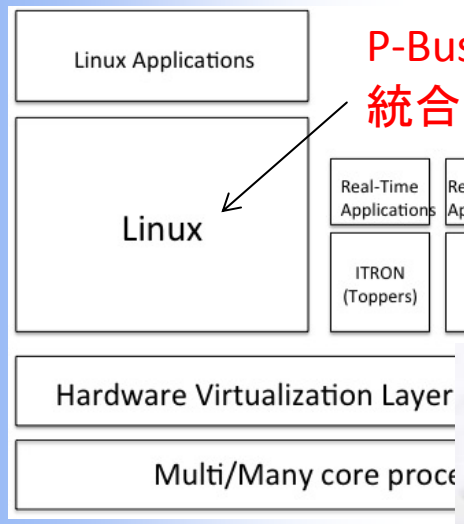
- モニタリングサービスはシステム全体の整合性　管理をおこなう汎用的なフレームワーク.
  - 各OSの振る舞いやメモリの状態に関する一貫性に関するルールを定義
    - 振る舞い: OSのハング，リアルタイム制約の崩壊
    - メモリの状態:データ構造の一貫性
    - PDCAによる新しい定義の追加
      - 整合性毎に定義する
      - ターゲットに依存したものの追加
      - 新しい脅威に対応する追加

・信頼出来ないP-Componentが挿入された場合にモジュールを消去する
・攻撃により隠蔽されたプロセスを消去する

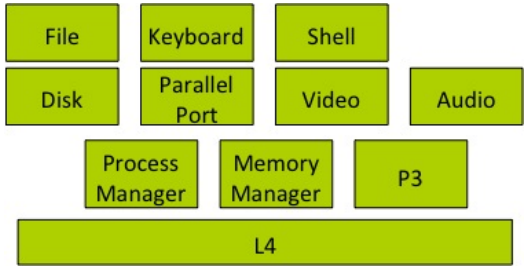


研究の現状

対応OS: Linux, Toppers, L4
最小限の変更により仮想化層上で動作可能

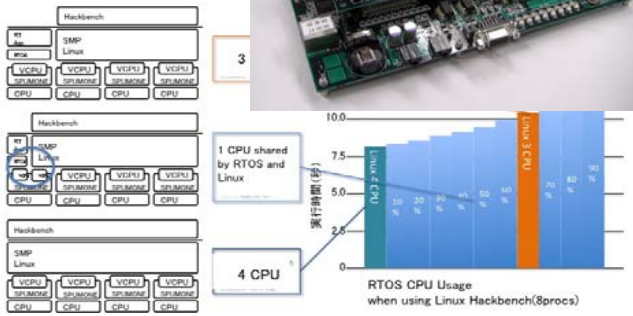


P-Bus対応Linuxとの統合も終了

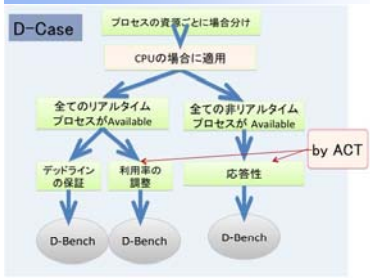


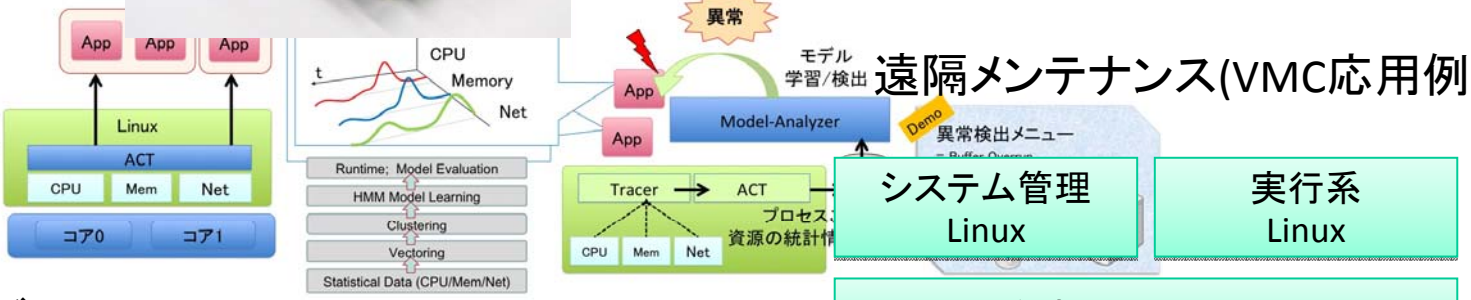
EOS
プロアクティブ再起動

マルチコア仮想化プロトタイプ（VMC）



VMC/VMO連携





遠隔メンテナンス(VMC応用例)

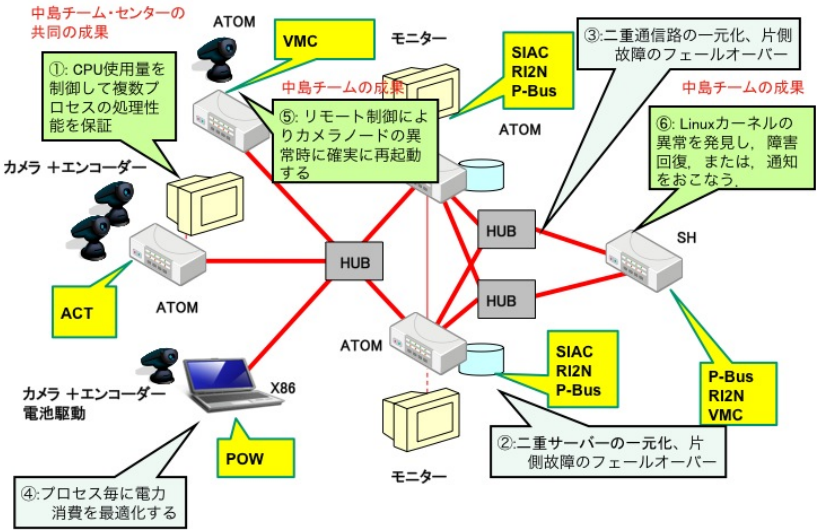
アカウンティング／ロギングシステム（ACT）

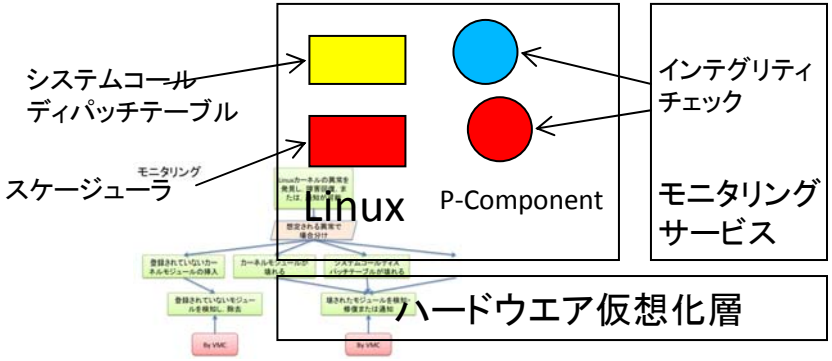
e-Societyの成果をセンターに移管，
また，ロギング機能等をセンターと共同で追加

シングルコア版の仮想化層はP-BUS対応Linuxとインテグレートしてセンターに技術移管予定（10月）

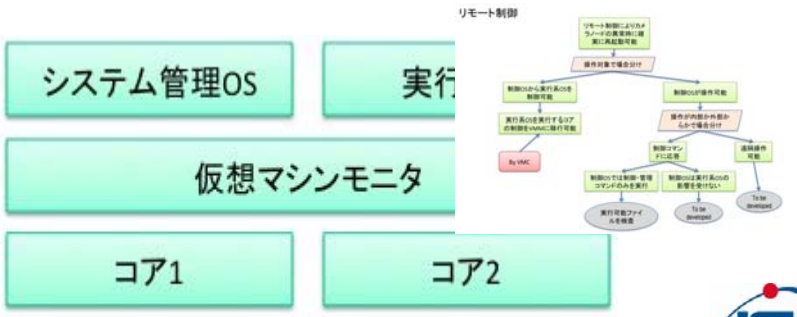
デモストレーション

- モニタリングデモ（統合デモ，個別デモ）
  - カーネル内のデータ構造の異常はシステムを正常に動作することを保証しない．
  - P-Component, システムコールディスパッチャ，スケジューラの異常を発見してシステムの自動修復をおこなう．
    - 信頼出来ないP-Componentは挿入出来ない
    - 古いバージョンのP-Componentは挿入出来ない．
    - 攻撃により隠蔽されたプロセスを自動的に削除する
    - 攻撃により変更されたシステムコールテーブルを修復する．
- リモート制御デモ（統合デモ，個別デモ）
  - OSのバグによりアプリケーション側のサービスが止まった時に，ユーザがクライアント・デバイスのところまで出向いてリセットしないといけない．
  - リモート管理サーバから異常を発見したときに再起動することによりサービスを回復する．
    - サービスを実行するOSがハングした場合でも異常を発生したOSの再起動が可能となる．
- マルチコア対応仮想化層(VMC), EOS(高信頼RTOS), アカウンティング・ロギングシステム
  - 個別デモで紹介





遠隔監視　　モニタリングデモ



重要な成果の概要

- 主な査読付き論文（詳細は最後のスライド）
  - ″SIGMA System: A Multi-OS Environment for Embedded Systems″, Journal of Signal Processing Systems(Accpeted 2008)
  - ``A Framework for Self-healing Device Drivers'', Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems, 2008.
  - ″A Lightweight Anomaly Detection System for Information Appliances″ , 12th IEEE International Symposium on Object/component/service-oriented Real-time distributed Computing (ISORC 2009) Tokyo, Japan, March 17-20, 2009.
  - ``Building a Self-Healing Embedded System in a Multi-OS Environment'', In the proceedings of the 24th Annual ACM Symposium on Applied Computing, 2009.
  - ″A Study on Asymmetric Operating Systems on Symmetric Multiprocessors″, In Proceedings of the 2007 IFIP International Conference on Embedded and Ubiquitous Computing, 2007(Best Paper Award).
  - ″Constructing MultiOS Platform with Minimal Engineering Cost″, International Embedded Systems Symposium 2009, 2009,
- 主催した国際会議やワークショップ
  - **12th IEEE International Symposium on Object/component/service-oriented Real-time distributed Computing (ISORC 2009)**
  - **First International Workshop on Software Technologies for Future Dependable Distributed Systems (STFSSD 2009)**
- DEOSの成果のアピール
  - 招待講演
    - **STARCシンポジウム200「「ディペンダブル組込みOSの挑戦」**
    - **CEATEC講演(10月)「マルチコア仮想化」**
    - **The Second International Workshop on Wireless Sensor Networks and Embedded Systems Research, “An Operating System for Future Multi/many-cores Information Appliances”**
    - 台湾大学，ウオータルー大学での講義
  - 実用化に向けた検討
    - 各社との共同研究や議論(ルネサス，富士通，東芝，**NEC,** パナソニック)を通した仮想化技術への要求事項の抽出
- VMMベンチマーク（進行中）
  - 推進委員と協力しながらメトリクスを明らかにしている
- DEOSセンターとの連携
  - アカウンティングシステム，ロギングシステムに関する研究
  - シングルコアバージョンの**CD**パッケージ化（10月）

結論

- 中島チームがDEOSプロジェクト内で開発している仮想化技術は以下の用な付加価値を従来のOSに追加する.
  - Linuxの異常時でもモニタリング機能を実行可能とすることによりシステム全体の信頼性を向上する.
    - モニタリングサービスをLinuxから分離し冗長化することにより信頼性を向上する.
  - 既存のRTOS上の資産を容易にDEOSシステムに統合することが可能となる.
    - マルチコアプロセッサ上で複数のOSを同時に利用可能とする.
  - 将来のマルチコアの進展に容易に追従することを可能とする.
    - 高度なパワーマネジメントや異機種性への対応を容易に実現する.
- 今後の課題
  - 仮想化技術プロトタイプシステムの性能・機能向上
  - リスク管理手法と仮想化技術の関連の更なる検討
  - 動的マルチコア管理技術のポリシに関する検討

査読付き発表論文, 招待講演リスト

原著論文発表

1. Yuki Kinebuchi, Hidenari Koshimae and Tatsuo Nakajima, Constructing Machine Emulator on Portable Microkernel, The 22nd Annual ACM Symposium on Applied Computing Seoul, pp.1197-1198, 韓国, 2007年3月
2. Yu Murata Wataru Kanda Tatsuo Nakajima, “A Study on Asymmetric Operating Systems on Symmetric Multiprocessors”, In Proceedings of the 2007 IFIP International Conference on Embedded and Ubiquitous Computing, 2007(Best Paper Award), pp.182-195, 台湾, 2007年12月
3. Yuki Kinebuchi, Midori Sugaya, Shuichi Oikawa, Tatsuo Nakajima, ``Task Grain Scheduling for Hypervisor-Based Embedded System'', In Proceeding of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC-08), pp.190-197, 中国, 2008年9月
4. Shingo Aoyagi, Shuichi Oikawa: IXIV VMM: A VMM on 2-Level Ring Architecture. IEEE 8th International Conference on Computer and Information Technology, pp. 533-538, オーストラリア, 2008年7月
5. Sun Lei, Tatsuo Nakjaima, ``A Lightweight Kernel Objects Monitoring Infrastructure for Embedded Systems'', In Proceedings of The 14th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2008), pp.55-60, 台湾, 2008年8月
6. Hiroo Ishikawa, Alexandre Courbot, Tatsuo Nakajima, ``A Framework for Self-healing Device Drivers'', Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems, pp. 277-286, イタリア, 2008年10月
7. Lei Sun, Tatsuo Nakajima, ``A Lightweight Detection and Recovery Infrastructure of Kernel Objects for Embedded Systems'', In Proceedings of The 2008 International Conference On Embedded and Ubiquitous Computing(EUC 2008), pp. 136-143, 中国, 2008年12月
8. Wataru Kanda, Yuki Kinebuchi, Yu Yumura, Tatsuo Nakajima, ``SPUMONE: LightWeight CPU Virtualization Layer for Embedded Systems'', In Proceedings of The 2008 International Conference On Embedded and Ubiquitous Computing(EUC 2008), pp.144-151, 中国, 2008年12月
9. Tomohiro Katori, Lei Sun, Dennis Nilsson, Tatsuo Nakajima, ``Building a Self-Healing Embedded System in a Multi-OS Environment'', In the proceedings of the 24th Annual ACM Symposium on Applied Computing, pp.293-29, アメリカ, 2009年3月
10. Midori Sugaya, Yuki Ohno, Andrej van der Zee and Tatsuo Nakajima, “A Lightweight Anomaly Detection System for Information Appliances” , 12th IEEE International Symposium on Object/component/service-oriented Real-time distributed Computing (ISORC 2009), pp.257-266, 日本, 2009年3月
11. Lei Sun, Dennis K. Nilsson, Tomohiro Katori and Tatsuo Nakajima, “Online Self-Healing Support for Embedded Systems” , 12th IEEE International Symposium on Object/component/service-oriented Real-time distributed Computing (ISORC 2009), pp.283-287, 日本, 2009年3月
12. Yutaka Ishikawa, Hajime Fujita, Toshiyuki Maeda,Midori Sugaya, Mitsuhisa Sato, Toshihiro Hanawa, Yuki Kinebuchi, Tatsuo Nakajima, Jin Nakazawa, and Hideyuki Tokuda “Towards an Open Dependable Operating System”, 12th IEEE International Symposium on Object/component/service-oriented Real-time distributed Computing (ISORC 2009), pp.20-27, 日本, 2009年3月
13. Ki-duk Kwon, Midori Sugaya and Tatsuo Nakajima, Analysis of Embedded Kernel using Kernel Analysis System, The 6th International Conference on Embedded Software and Systems, pp. 417-422, 中国, 2009年5月
14. Yuki Kinebuchi, Kazuo Makijima, Takushi Morita, Midori Sugaya, Tatsuo Nakajima, “CONSTRUCTING MULTI-OS PLATFORM WITH MINIMAL ENGINEERING COST”, International Embedded Systems Symposium 2009, 採録決定, ドイツ, 2009年9月
15. Andrej van der Zee, Alexandre Courbot, Tatsuo Nakajima, “mBrace: Action-based Performance Monitoring of Multi-Tier Web Applications”, In Proceedings of The 2009 International Conference On Embedded and Ubiquitous Computing(EUC 2009), 採録決定, 2009年8月
16. Wataru Kanda, Yu Murata and Tatsuo Nakajima, "SIGMA System: A Multi-OS Environment for Embedded Systems", Journal of Signal Processing Systems, 採録決定, 2009年

招待講演・論文

1. Tatsuo Nakajima, Software Infrastructure for Next Generation Computing Environments, Colloquium, Rutgers University, 米国, 2007年12月
2. 「ディペンダブル組込みOSの挑戦」STARCフォーラム/シンポジウム2008, 日本, 2008年7月
3. Tatsuo Nakajima, et. al., ``An Operating System Architecture for Future Information Appliances'', In Proceedings of IFIP International Workshop on Software Technologies for Future Embedded and Ubiquitous Systems, イタリア, 2008年10月
4. Yutaka Ishikawa, Hajime Fujita, Toshiyuki Maeda,Midori Sugaya, Mitsuhisa Sato, Toshihiro Hanawa, Yuki Kinebuchi, Tatsuo Nakajima, Jin Nakazawa, and Hideyuki Tokuda “Towards an Open Dependable Operating System”, 12th IEEE International Symposium on Object/component/service-oriented Real-time distributed Computing (ISORC 2009) 日本, 2009年3月
5. Tatsuo Nakajima, An Operating System for Future Multi/many-cores Information Appliances, The Second International Workshop on Wireless Sensor Networks and Embedded Systems Research, 韓国, 2009年6月