

# URLフィルタリングアクセラレータの開発

## Development of a URL filtering accelerator

米田 友洋 · 佐藤哲朗(株 ノディック) · 河口文法(株 ノディック) · 和田正伸(株 ノディック)  
Tomohiro YONEDA, Tetsuro SATO(Nodc inc.), Fuminori KAWAGUCHI(Nodc inc.), Masanobu WADA(Nodc inc.)

### どんな研究?

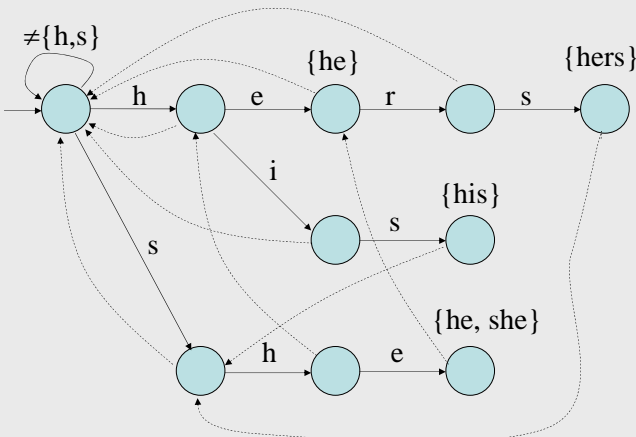
近年違法、有害サイトが問題となり、URLフィルタリングやコンテンツのキーワードフィルタリング等が重要になってきました。禁止するURL数やキーワード数が巨大でもリアルタイムでフィルタリングできるハードウェアの研究です。

### 何が特徴?

外部メモリ上に巨大なオートマトンを書き込み、これでパターンマッチをおこないます。スループットはメモリアクセスに律速されますが、非同期式パイプラインは途切れないメモリアクセスを実現できるので最大限のメモリスループット、パターンマッチスループットを引き出します。

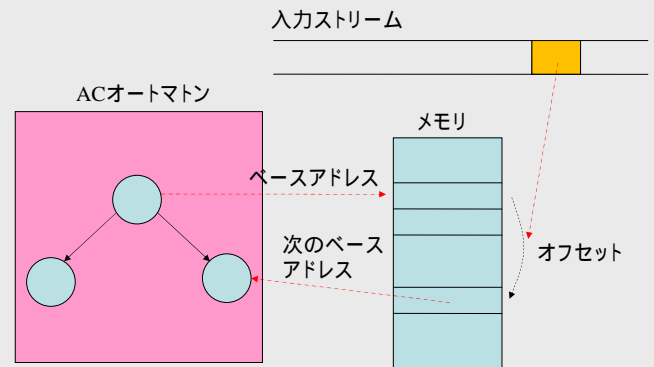
### パターンマッチングの原理: Aho Corasick法

- キーワード照合文字列集合の例 : { he, she, his, hers }
- 対応するACオートマトン



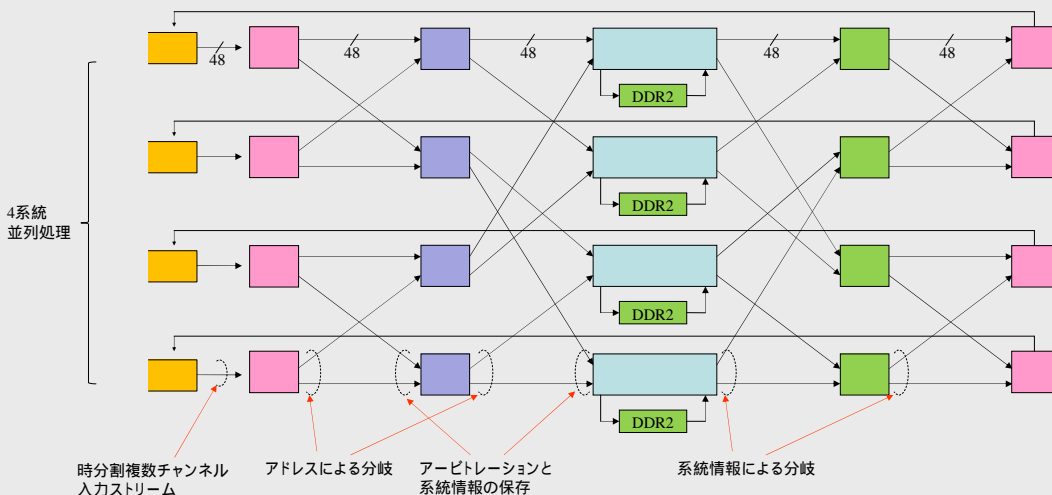
- ストリームに対するマルチパターン検索が可能

### アクセラレータのアイデア



- 外部メモリ上にACオートマトンを展開
  - パターンがハードワイヤードされないため更新が容易
  - 巨大なパターンを搭載可能
- 並列化・パイプライン化によるマルチストリーム処理が容易
- そのほかの応用(正規表現フィルタリング, XMLパーサ等)にも適用可

### 非同期式パイプラインによるデータフロー構造の実現



- 非同期式による利点
  - 複雑かつ不均一なパケットを早いもの順に流すパイプラインができるので、パケットはDDR2メモリの前で自然に待ち行列となる。このためメモリアクセスは均一で途切れることがない。
- FPGA(Spartan-3)による実装
  - 外部DDR2メモリ使用
  - 150万URLを含むオートマトン(5000万URLまで可能)
  - 約1Gbpsのスループット
- 今後の課題
  - 正規表現フィルタリングへの展開
  - ICAPサーバ化
  - 更に高スループット化