ネットワークセキュリティのための CAMによる文字列検索システム

The String Search System Using CAM for Network Security

村田 健二† 原 悠希† 中西 衛‡ 小倉 武†
Kenji MURATA† Yuuki HARA† Mamoru NAKANISHI‡ Takeshi OGURA†
†立命館大学大学院理工学研究科 〒525-8577 滋賀県草津市野路東 1-1-1
‡NTTマイクロシステムインテグレーション研究所 〒243-0198 神奈川県厚木市森の里若宮 3-1

1. はじめに

インターネットが世界中に普及し常時接続が一般的となった今，コンピュータウイルスやワーム，不正アクセスといった悪意ある攻撃の脅威はますます高くなっている．そのため，各種セキュリティリスクに対するセキュリティの確保は必須となっている．

一般的にセキュリティ処理の中核として文字列検索が行なわれている[1]が，プロセッサによるソフトウェア処理では検索処理が追いつかず，最悪の場合には攻撃のある通信を見落としてしまう可能性がある．そのため基幹ネットワーク等では専用ハードウェアが用いられることが多い．

本論文ではハードウェア化案として，CAM の並列処理を生かした高速な文字列検索システムを提案する．そして CAM のインストラクションレベルのシミュレータにより特性を評価し，実現可能であることを確認した．

2. 使用するデバイス

2.1 CAM

CAM(Content Addressable Memory)は日本語で連想メモリよばれる機能メモリである．RAM など通常のメモリだとアドレスを指定してデータにアクセスする，それに対して CAM はデータに直接アクセスして検索，書き込みをすることができる．また，検索処理は全ワードが並列動作することで高速動作を実現している．

2.2 BCAM と TCAM

一般の CAM は BCAM(Binary CAM)であり 0,1 の二値しか記憶することができない．これに対してネットワークセキュリティ機器でよく使われる CAM は TCAM(Ternary CAM)であり，0,1，*(ドントケア)の三値を記憶できるように拡張したものである．これにより完全一致でなくとも検出が可能なため，より自由度の高い検索ができるようになる．一つの検索キーに対し，確定記号のみのワードとドントケアを含むワードとが複数ヒットした場合がある．この場合，TCAM では LPM と言う手法により選択分離されている．LPM は，確定記号がより長いワードを選択する機能である．

しかし BCAM に比べ，記憶セルの大きさが約二倍必要であり，さらに LPM を実現するためにプレソーティングをする別回路が必要になるなどハードウェア量が大きくなってしまうという問題がある．

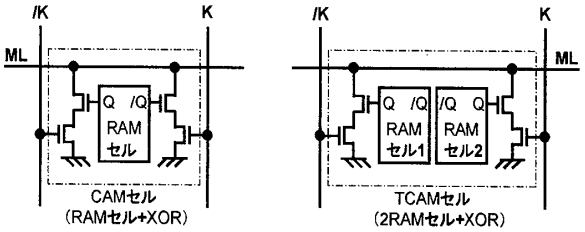

図 1. BCAM と TCAM の記憶セルの構造

3. 検索アルゴリズム

3.1 データの格納法

データは一文字をASCII8 ビット+1 ビットの計 9 ビットで表す．末尾の 1 ビットで確定記号，ドントケアの判別をしており末尾のビットが 0 なら確定記号，1 ならドントケアを表す．さらに各ワードに検索対象を判定するビット(検索対象フラグ)，ドントケア一致を判定するビット(ドントケアフラグ)，作業用のビットの計 3 ビットを付加している[2]．

CAM 幅より短いパターンは余りをドントケアで埋め，長いパターンは CAM 幅に分割して格納する．このとき短いパターンを作らないように前の文字を重複させて CAM 幅に調節する．

CAMに格納したデータはユニークなものとしてそれぞれにインデックスをつける．もとのシグネチャが復元できるように対応をテーブルに記憶する．パターンの種類を記憶したテーブルをPT，パターンの組み合わせを記憶したものをMTとする[3]．

3.2 ドントケア検索と LPM

CAM の検索は，記号直列に行い，一文字に二度の検索を行うことでドントケア検索を可能にしている．1st 検索はドントケア，2nd 検索は確定記号を検索キーとして行う．最初の文字は全ワードに検索を行う．1st 検索でヒットした場合は全ワードのドントケアフラグを 0 にしてヒットしたワードのドントケアフラグを 1 にする．2nd 検索後，一致しなかったワードの検索対象フラグを 0 にする．以降は検索対象フラグが 1 のワードに対して検索を続けていき，最後の文字の 2nd 検索でヒットがあったアドレスを結果として返す．2nd 検索でヒットがなかった場合はドントケアフラグが 1 のアドレスを結果として返す．全ワードで検索対象フラグ，ドントケアフラグが 0 になったとき不一致として検索を終了する．

この手法ではワードの並びに関係なく LPM を実現することができる．そのため他の手法のようなプレソーティ

†Department of VLSI System Design, Ritsumeikan Univrsity
1-1-1 Noji-higashi, Kusatsu-shi, Shiga, 525-8577 Japan
‡NTT Microsystem Integration Laboratries
3-1 Morinosato, wakamiya, Atsugi-shi, Kanagawa, 243-0198 Japan

ングは必要なく，シグネチャの追加も容易となる．

3.3 各種シグネチャへの対応

テーブル PHL を用意してパターンが見つかった場所，見つかったパターンのインデックスを記憶する．まずは検索データを CAM 幅で切り出し，CAM 内のデータと照合する．一致しなければ 1 文字シフトして検索を繰り返す．一致すれば PT で種類を確認し，短いパターンのとき報告する．先頭のパターンの場合は PHL に記憶，途中のパターンのときは MT により，PHL のデータと組み合わせて有効なパターンができるか確認する．組み合わせが見つかった場合，元のシグネチャが復元できたら報告する．新たな先頭を形成したらそのインデックスを PHL に記憶する．検索範囲を越えている PHL を削除する．それから 1 文字シフトして検索を繰り返す．

4. システムの特性

4.1 評価環境

CAM のインストラクションレベルのシミュレータを構築し，動作サイクル数で特性を評価する．使用するシグネチャは Snort のルールから取り出した 200 個である．図 2 に例を示す．

```
#1. content:"Bad command or filename";
#2. content:"1 file|28|s|29| copied";
#3. content:"HTTP/1.1 403";
#4. content:"uid="; content:" gid="; distance:15;
#5. content:"MicrosoftWindows"; content:"|28|C|29|Copyright 1985-
"; distance:10; content:"Microsoft Corp."; distance:10;
```

図 2. 使用したシグネチャの例

単一のパターンで構成されているもの(#1,2,3)だけではなく，複数のパターンが離れて分布しているような複雑な構成のシグネチャ(#4,5)も対応している．

4.2 シグネチャ数の依存

ソフトウェアでの検索だとシグネチャ数の増加により比例的に処理時間が増加してしまう．しかし CAM は全ワードを並列に検索できることからワード方向の増加に影響されにくいと予想できる．

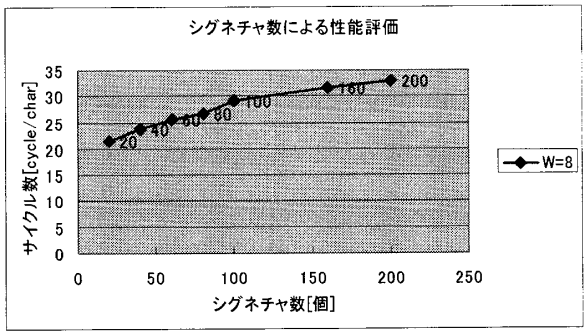

図 3. シグネチャ数による評価

シミュレーション結果は図 3 のようになりシグネチャの増加につれてサイクル数も増加した．しかし，傾きは小さく影響を受けにくいといえる．

4.3 CAM 幅とワード数

CAM の記憶容量について考察する．最も影響するパラメータとして CAM 幅が考えられる．シグネチャを CAM 幅で区切って格納していくため CAM 幅が短くなるとワード数が増加する．増加の程度はシグネチャの長さ分布に依存する．今回使用した 200 個の場合は表 1 のようになった．

表 1. CAM 幅とワード数

CAM 幅	ワード数
4	538
6	429
8	365
12	296
16	263

使用を前提としている CAM は 64 ビット×16 k ワードで構成されている．表 1 の関係を前提とすると CAM 幅 4 の場合で 6000 程度，CAM 幅 6 の場合だと 7500 程度のシグネチャを 1 チップに格納できる．

5. まとめと今後の予定

本論文では BCAM による文字列検索システムを提案し，特性を評価し実現可能であることを示した．これにより 1 つの文字につき 1 ビット付加するだけでドントケア検索が可能になる，つまり TCAM を用いることなく BCAM により処理を実現することでハードウェアコストや消費電力の問題を解決できる．今後は図 4 で示す FPGA 搭載 CAM ボードを用いて本システムを実装し，提案方式の有効性を検証する予定である．

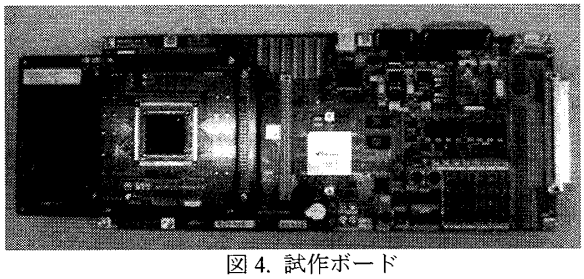

図 4. 試作ボード

参考文献

[1]榑林亮介，片山勝，山中直明，塩本公平，“ハッシュ型オートマトンによる文字列検索の並列化手法”，信学技報，CS2003-166, pp4-54, Mar.2004.

[2]山野達彦，喜多健雄，村田健二，中西衛，小倉武，“ネットワークセキュリティのための CAM による文字列検索システムの検討”，FIIS07, No.215, Ooita, Japan, Jun.2007.

[3]Fang Yu, Randy H.Katz, T.V.Lakshman, “Gigabit Rate Packet Pattern-Matching Using TCAM”, Proceedings of the 12th IEEE International Conference on Network Protocols 2004. (ICNP'04), pp.174-183, Berlin, Germany, Oct.2004.