センサーからの情報に基づいたシグネチャ自動生成システムのモデル化と基礎検討

小林　武史[†1]　飯田　勝吉[†2]
中村　豊[†3]　池永　全志[†4]

A Preliminary Study and Modeling of Signature Generation System based on Distributed Sensors

Takefumi Kobayashi,[†1] Katsuyoshi Iida,[†2] Yutaka Nakamura[†3]
and Takeshi Ikenaga[†4]

1. はじめに

近年，ネットワークの大容量化に伴い，ネットワーク障害が全体に及ぼす影響が大きくなっている．こうした状況から，障害を早期に防止することが求められる．しかし，インターネットの進化とともに異常のタイプは多様化し，変化のスピードが早くなっているので，人の手だけではとても対応することができない．その中で特に，自己増殖するものの被害が深刻になっている．たとえば，ワームはホスト間を感染するので，広がる速度が速く，一度広がったワームに対応するのは難しい．このように，インターネットを介して感染していく異常は広がるスピードが早く，ネットワーク全体に大きな被害をもたらす可能性がある．

そこで，本研究では感染する異常に対して早急に対応することを目的とし，センサーからの情報に基づいたシグネチャ自動生成システムについて検討をする (図 1)．まず，複数のセンサーがそれぞれ異常検出を行い，情報をデータセンターに送る．データセンターはセンサーからの情報に基づいて異常検出を行う．さらに，データセンターは検出した異常に対してシグネチャの生成を行う．このように異常検出とシグネチャ生成という 2 つのステップから成る．今回は，実際にシステムを作るための準備段階として性能を調査するために異常検出手法のモデル化を行う．

本手法では，各センサーは独立して異常検出を行い，データセンターが複数のセンサーから情報を得て感染する異常を早期に検出することができる．2. 章では既存の手法に関していくつかを挙げる．3. 章では提案手法のモデル化を行う．4. 章ではシミュレーションによる提案モデルの評価の結果の例を示す．

†1 東京工業大学工学部 開発システム工学科
†2 東京工業大学　学術国際情報センター
†3 九州工業大学　情報科学センター
†4 九州工業大学大学院　工学研究科

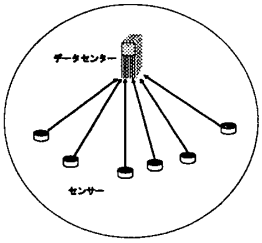

図 1　システム概要

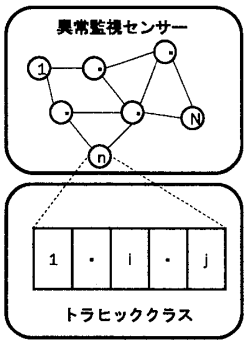

図 2　モデル化の概要

2. 既存の手法

1) では複数点によって異常検出を行い，異常の原因を特定するシステムを提案している．ここでは，異常のタイプを識別し，検出された異常を種類別に分類することができる．

本研究では，さらに異常が伝搬するモデルを作り，感染する異常を検出することを目標とする．また，シグネチャ生成という次の目標を見据えて異常検出のモデルを作る．

3. 異常検出方法のモデル化

提案手法は，2 つのステップからなる．まず，各センサーが独立して異常検出をする．次に，データセンターがそれらのセンサーからの情報に基づき，感染するものについて異常検出を行う．以下で，それぞれについて説明する．

3.1 センサーの動作

初めに，各センサーで行う異常検出について説明する．各センサーに到着するトラヒックは，j 種類のトラヒッククラスに分けられ，トラヒッククラスごとに到着する (図 2)．ここで，トラヒッククラスの分類方法は，IP アドレス，ポート番号，TCP フラグを用いる方法などが考えられる．各センサーでは，トラヒッククラスごとのトラヒックの比率を計測している．タイム

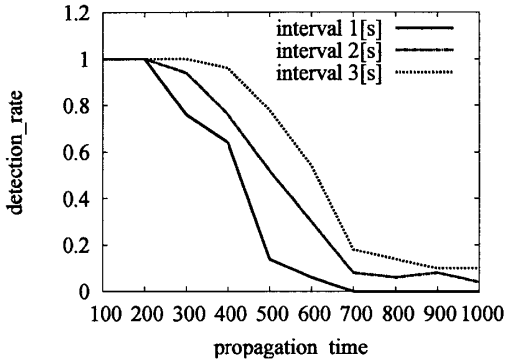

図 3 検出精度 (持続的に異常トラヒックが流れる場合)

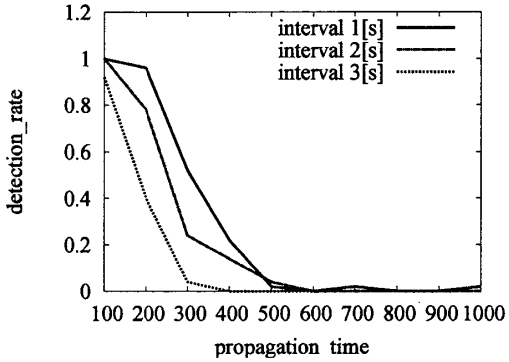

図 4 検出精度 (瞬間的に異常トラヒックが流れる場合)

スロット k におけるトラヒッククラス i のトラヒック量の計測値を $p_i(\sum_{i=1}^{j} p_i = 1)$ とし，トラヒッククラスごとのトラヒック量の時間的な変動に着目する．具体的には，以下の移動平均を用いる．

$$(\hat{p_i})_k = (1-\alpha)(\hat{p_i})_{k-1} + \alpha(p_i)_k (0 \leq \alpha \leq 1)$$

ここで，現在のラウンドの計測値から現在のラウンドの移動平均値を引いた値 $(p_i)_k - \hat{p_i}_k$ に着目し，この差がある閾値よりも大きくなった場合に該当トラヒッククラスにおいて異常が発生したと判定する．ここで，閾値の設定には信頼区間 c %を利用する．また，トラヒック量の一時的な増加による誤検出を排除するために，3 タイムスロット連続で閾値を越えた場合にそのセンサーで異常が発生したと判断し，そのセンサーは該当トラヒッククラスに対してアラームフラグを立てる．フラグは，有効期間 v だけ立て，v を過ぎたら消す．

3.2 データセンターの動作

データセンターは，N 個のセンサーからの情報を統合する役割を担う．具体的には，どのセンサーにアラームフラグが立っているのかを把握し，同種類のフラグの数が閾値 N_a を超えたとき，該当するトラヒッククラスに異常が発生したと判定する．

表 1 評価パラメータ

名称	数値
センサー数 (N)	30
各センサーの速度	4Mb/s
パケットサイズ	500Byte
トラヒッククラス数 (j)	4
移動平均の重み (α)	0. 001
信頼区間 (c)	99[%]
有効期間 (v)	2000[s]
センサー数閾値 (N_a)	10
サンプリング間隔	1, 2, 3[s]
異常トラヒック増加率	0. 2, 1
異常トラヒック継続時間	∞, 10[s]
計測時間	10000[s]

表 2 各トラヒッククラスの比率

p_1	p_2	p_3	p_4
0.2	0.4	0.1	0.3

4. シミュレーション

本章では，シミュレーションにより 3 章で作ったモデルの数値結果の例を示す．ここでは，異常は異常伝搬係数 τ に従い感染していくとする．異常伝搬係数とは，異常が他のセンサーに感染する指数分布に従う平均時間である．0[s] にトラヒッククラス 2 に対して異常トラヒックが流れ込んだ場合，τ を動かした場合の，検出精度と検出時間を測定した (ただし，誌面の都合より検出時間については割愛する)．検出精度とは，各異常伝搬係数に対して検出時間以内にうまく検出できた割合を表す．また，検出時間とは 0[s] に異常トラヒックが流れてから検出するまでの時間平均時間を表す．

ここでは，以下の 2 つの場合について考察する．持続的に異常トラヒックが流れ込む場合と瞬間的に大きな異常トラヒックが流れ込む場合である．それぞれの場合に対する検出精度を図 3，図 4 に示す．

ここでは，前者の場合はサンプリング間隔が大きいほうが検出精度がよく，逆に，後者の場合は，サンプリング間隔が小さいほうが検出精度がよくなっている．このような傾向が見られるのは，サンプリング間隔が大きいほうが分散が小さく持続的な異常に対して安定して対応できるためである．また，サンプリング間隔が小さいほうが一時的な変化を見つけやすいので，検出精度が高くなっている．

5. おわりに

本稿では，複数のセンサーとデータセンターによる異常検出機構のモデル化を行い，実際にシミュレーションを行った．その結果，持続的に異常トラヒックが流れる場合はサンプリング間隔が大きいほうが検出精度がよく，瞬間的に大きな異常トラヒックが流れる場合，サンプリング間隔が小さいほうが検出精度がよくなることが分かった．今後の課題として，今回発見した性質を基に，提案モデルについてさらに考察を深めて行く．

謝　辞

本研究は，科研費特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」の援助を受けています．

参考文献

1) 中村信之，中井敏久，「複数プローブによる異常トラフィック検知システム」，情報処理学会，研究報告，2006-DPS-126，2006 年 3 月