Vol.2017-DPS-170 No.28
Vol.2017-CSEC-76 No.28
2017/3/3

ログの順序パターンに基づく異常検知手法の提案とCANのログへの適用

栗田 萌[1,a)]　渡辺 俊貴[1]　山野 悟[1]

概要：自動車内部の制御ネットワークと、インターネットのような広域ネットワークの双方に接続されるシステムが普及することによって、自動車の利便性が増す一方で、制御ネットワークの内部への不正なアクセスのリスクが懸念されている。このような背景の下、自動車の制御ネットワークとして広く普及している CAN(Controller Area Network) の通信を安全にするための手法が研究されてきた。CAN の通信はその正常状態が一定の周期的性質を持つ場合が多く、従来研究において時間間隔の正常状態に注目した異常検知手法などが提案されている。本論文では、メッセージの周期性に加えてメッセージの出現順序の特徴を CAN のログより見いだし、その正常状態を抽出する手法を提案し、その結果を示す。

Method for Anomaly Detection Based on Sequential Pattern of Log and Its Application to CAN Log

Moyuru Kurita[1,a)]　Toshiki Watanabe[1]　Satoru Yamano[1]

Abstract: While spread of the system which connects control networks inside vehicles and broad networks such as Internet enhances convenience of vehicles, risks for malicious access to control network are concerned. Therefore, method for realizing secure and safe communication in CAN(Controller Area Network) which is the wide spread control network in vehicles has been studied. Since CAN communication often has periodic properties in its normal state, anomaly detection methods using its time period are proposed in conventional studies. In this paper, we propose a method for extracting property of normal state by focusing on periodic sequence of messages and show its results.

1. はじめに

近年では IoT(Internet of Things) の発達により、パソコンやサーバ、プリンタ等の従来の IT 関連機器だけでなく、自動車や監視カメラ、制御装置などがインターネットに接続される機会が増えている。これにより機器間の情報伝達が迅速化するとともに、より高度な機能を追加することが可能となってきている。

その一方で、今まで内部に存在するネットワークのみで情報伝達を行っていた機器が外部と接続を持つことにより、セキュリティの問題が発生するリスクが高まっている。特に自動車や制御装置に対する攻撃は人命に関わる事故に直結する可能性を含んでいるため、侵入検知や防止などの対策が急務となっている。

このような背景があり、近年では CAN(Control Area Network) をセキュリティの問題から守る研究が活発になされるようになってきている。そのためのアプローチとして、CAN のメッセージが正常時の挙動を学習し、それと異なった挙動が見られた場合に異常として検知する手法が考えられる。例えば、CAN の場合流れるメッセージが一定の時間周期を持つ場合が多く、これが乱されたときに異常と判断する手法が考えられる。しかしこの手法の場合、時間周期を保ちつつメッセージの順序が変わるような異常検知を行うことはできない。そこで本論文では、自動車の CAN のログを例に、ネットワーク内のメッセージの順序

[1] 日本電気株式会社　中央研究所
211-8666 川崎市中原区下沼部 1753 番地
NEC Corporation, Central Research Laboratories
1753 Shimonumabe Nakahara-ku Kawasaki,
Kanagawa 211-8666, Japan.
[a)] m-kurita@ap.jp.nec.com

に関して特定の法則が見出されることに注目し、正常状態の抽出および異常の検出を行うシステムを提案する。

2. 先行研究

2.1 CAN の攻撃手法に関する先行研究

車載ネットワークに対する攻撃は 2010 年頃から盛んに行われるようになり、Hoppe らによる ECU への攻撃実験 [1] や、診断端子 (OBDII) から CAN バスにデータを不正に挿入し ECU の不正制御を行った Kosher らの研究 [2]、さらに、それを無線通信で実現する Checkoway らの研究がある [3]。また、Miller らは 2013 年に車内からの攻撃手法として、診断ポートを介して運転支援機能を悪用し、ステアリングやブレーキの不正制御を実証している [4]。さらに、2015 年にヘッドユニットのファームウェアを、ヘッドユニットのネットワーク接続機能の脆弱性をついて書き換え、そのファームウェアを介して遠隔地から CAN メッセージを送信することで、車両を不正に制御できることを実証した [5]。この結果、100 万台を超える自動車がフィアットクライスラー社からリコールされる事態に発展した。

2.2 CAN の保護手法に関する先行研究

CAN を保護する手法として代表的なものは、本論文で提案するような異常検知手法を用いたものや暗号技術を用いる方法がある。暗号技術を用いるものでは、メッセージに Message Authentication Code(MAC) を付加することでメッセージの信頼性を高める手法が複数提案されている [6][7] ほか、AUTOSAR(AUTomotive Open System ARchitectur) において標準化技術としての検討もなされている。

また、機械学習の手法を用いてメッセージの正常状態を学習し、それとの比較を行うことで異常検知を行う手法も提案されている。CAN のメッセージは周期的に送信される場合が多く、送信間隔を特徴量とする手法 [8][9] や、メッセージ内部の変数値を特徴量とする手法 [10] が提案されている。この他にメッセージの頻度に関する情報エントロピーを用いて異常を検知する手法が提案されている [11]。また、発信、停止、右左折などの運転者のふるまいが、CAN のメッセージに影響を与える順番を調べて、その順番を元に異常を検知する手法が提案されている [12]。

2.3 本論文が扱う課題

異常検知の手法の主流は正常な状態を何らかの手法でモデル化し、それと明確な差異が現れた場合に異常と見なす手法が一般的である。CAN に対する攻撃は研究としては進められているものの、実際に攻撃が行われたデータを元にして大規模な学習を行うことができるほど攻撃事例は存在していない。このような条件では、CAN のメッセージに関する正常状態を正しく抽出することは異常検知システムの性能を考える上で重要である。実際に CAN のログを分析したところ、正常状態で保持される順序が存在することが確かめられたため、この性質を利用することで異常検知の高精度化に役立つと考えられる。そこで、本論文では従来の CAN の正常状態抽出では扱われてこなかった、メッセージ一つ一つの順序に関する正常状態の抽出方法および、その性質に基づいた異常検知手法を提案する。

3. 提案手法

3.1 本論文で用いる CAN のログの性質

本節では CAN のログについての基本的な性質を説明する。CAN のログは以下の 3 つの要素を持つメッセージの列である。

- **タイムスタンプ**: メッセージを受信した時刻をミリ秒で記録したもの。
- **メッセージ ID**: 一つのメッセージは必ず一つの固有の番号を持つ。この番号はメッセージを送信した ECU の種類や、その ECU が送信するデータの種別を表している。以降、単に ID とした場合このメッセージ ID を指すものとする。
- **データ**: 整数や実数の値が格納されている。変数の次元や意味はメッセージ ID により様々である。

典型的な CAN メッセージのログは表 1 のようになっている。ここで、タイムスタンプは UNIX 時間で記されている。データについては、実際には多次元のものも含んでいるが、本論文ではタイムスタンプとメッセージ ID および、メッセージがログに現れる順序のみから抽出される正常状態を扱うため、簡単のため 1 次元のもののみ表記した。

表 1 代表的な CAN のログ

タイムスタンプ	メッセージ ID	データ
1458734342227	304	12642
1458734342227	734	800
1458734342227	853	64
1458734342232	374	12
1458734342233	384	11264
1458734342233	469	214
1458734342233	505	0
1458734342233	768	5
1458734342237	304	12899
1458734342237	730	800
⋮	⋮	⋮

また、このログに対して 3 つの性質が見られた。一つ目は、「多くの ID はその ID に固有の時間周期を持ち、その周期におおよそ沿った形でログに現れる」という性質である。これは、より具体的な例では「ID が 10 のメッセージは 20ms ごとにログに現れる」ということである。「多くの ID」と表現したのは、特定の時間周期を持たない ID も存

在するためである。

二つ目は、「この時間周期は分散を持つ」という性質である。これも、より具体的な例では「ID が 10 のメッセージが時刻 t にログに現れたら、次は t+20ms に現れることが多いが t+19ms や t+21ms の場合もある」というものである。我々が用いた CAN のログデータでは、時間周期の値が大きいほど分散も大きい傾向が見られた。時間周期が 10ms の ID を持つメッセージは前回の同一 ID を持つメッセージからおよそ 10±1ms の間に現れるが、時間周期が 500ms の ID を持つメッセージは 500±10ms の間に現れるという具合である。

三つ目は、「同一の時間周期を持ついくつかの ID の集合は、一定の順序を保ってログに出現する」という性質である。例えば ID10,11,12,13,14,15,16 のメッセージが 20ms の時間周期を持つ場合を考える。ログの中でこれらの ID を持つメッセージだけに注目すると、これらの ID のメッセージは同じ時間周期であることから、一定の時間内（例として数秒) は同一の順序でログに現れる (例:10,11,12,13,14,15,16,10,11,12,13,14,15,16,10.....) が、この順序は時間周期の分散によって変更するものであり、また CAN システムの起動条件などでも変わるため、長期的にログを調べたり、異なるログデータを調べた場合「一定の順序」が存在することは自明には存在しない。しかし、この中の一部の ID の集合は一定の順序に従う性質があった。例えば、ID10,11,12 を持つメッセージのみに注目すると (10,11,12,10,11,12,10,11,12,.....) という並びが常に保持されるような性質が見られた。

3.2 提案方式概要

前節で述べたような性質は我々が入手した CAN のログでは頻繁に確認された現象である。通常一つのログに対して、多数 (50 程度) の ID が存在し、それらの多くは時間周期によっていくつかのグループに分類することが可能であり、以下では一つのログを時間周期によって分類したものを考える。例えば時間周期 10ms のグループ、20ms のグループ…といったものであり、一つのグループに 2〜10 程度の ID が属している。実際、CAN のメッセージが時間的な周期性を持つことに注目し、異常検知を行う手法を提案した研究は複数存在する。

本論文では三つ目の性質に関連した CAN のログに見られたメッセージの順序について、自明でない法則性を見つける手法を提案するとともに、実際の CAN のログに対する適用結果を報告する。さらにこの性質を用いた異常検知手法についても提案する。この手法は、CAN のログだけでなく、正常時には定常的な動作を繰り返すシステムなどの異常検知に応用できると考えられる。

実際に、特定の周期の ID を持つメッセージだけを抽出したログの例を表 2 に示す。これらの ID はいずれも 40ms 間隔でログに現れる。順序に注目すると、この範囲内では ID660 と ID545 は順序の入れ替えが起こるが、それ以外の ID は順序関係を一定に保っている。このような性質を持つ理由は ECU の詳細な設定や、各 ECU の起動のタイミングなどによると考えられる。一方で CAN のメッセージが起動のタイミングなどによらず一定の順序を持つ ID の集合が存在する場合は、それは ECU の仕様を反映していると考えられ、異常検知の強化に利用できると考えられる。以下の節では、このような性質を抽出する具体的な手法を提案する。

表 2 特定の周期の ID を持つメッセージのみを抽出したログ

タイムスタンプ	メッセージ ID
1473777625026	57
1473777625039	597
1473777625053	660
1473777625053	545
1473777625053	661
1473777625066	57
1473777625079	597
1473777625093	660
1473777625093	545
1473777625093	661
1473777625106	57
1473777625119	597
1473777625133	545
1473777625133	660
1473777625133	661
⋮	⋮

3.3 一定順序集合の抽出手法

この節では、一定の順序集合を抽出するためのアルゴリズムを説明する。そのために、今回扱うログを扱う用語を定義する。はじめに $\mathcal{S} = \{s_1, s_2, \cdots, s_{|\mathcal{S}|}\}$ を ID の集合とする。これは、CAN のログに現れる ID の集合である。一般に CAN のログのメッセージ ID は任意の整数を取ることができるが、$[1, |\mathcal{S}|]$ までの整数に限定しても一般性を失わないため、以下ではそのようにする。そして、メッセージ列 $\mathcal{D}_{\mathcal{S}} = \{d_1, d_2, \cdots, d_{|\mathcal{D}_{\mathcal{S}}|}\}$ を ID と、その ID のタイムスタンプからなるメッセージ $d_i = (s(d_i), t_i)$ の列とする (ここで $s(d_i)$ は 1〜$|\mathcal{S}|$ までの整数である)。ここで、メッセージ列 $\mathcal{D}$ には以下の特徴があるとする。

- 一つの ID に注目すると、それは一定間隔に近い頻度で表れ、この間隔は全ての ID で等しい。すなわち、メッセージ列から ID が s_k のメッセージだけを取り出して構成した列 $\mathcal{D}_{\{s_k\}} = \{d_{k(1)}, d_{k(2)}, \cdots\}$ は

$$t_{k(j+1)} - t_{k(j)} \approx t_p \tag{1}$$

を満たす。

- 任意の $1 \leq i \leq |\mathcal{D}_\mathcal{S}|$ に対して以下のいずれかを満たす $1 \leq j \leq 2|\mathcal{S}|$ が存在する。
  - d_i と d_{i+j} の ID が等しい。
  - $i + j - 1 = |\mathcal{D}_\mathcal{S}|$

一つ目の性質は、前節で述べたとおりここで扱うログは一定の分散の範囲内で同一の時間周期を持つ ID の集合に分割されたものであるということを述べている。第二の性質は、第一の性質から自然に帰結されることであるが、一つの ID があらわれてから $2|\mathcal{S}|$ の間に同じ ID があらわれるか、ログが終端に達するということを述べている。また以下では 2 つ目の性質に関連して、i に対して上述した条件を満たす j を $j(i)$ と表す。

ここで、$\mathcal{S}$ の部分集合で、$\mathcal{D}_{\mathcal{S}'}$ が周期的となる $\mathcal{S}'$ を取り出す問題を考える。これは以下のアルゴリズムで可能である。M を $|\mathcal{S}| \times |\mathcal{S}| \times |\mathcal{S}|$ 次元の配列とし、その初期値を 0 とする。そして、1 から $|\mathcal{D}_\mathcal{S}|$ までの全ての i に対して以下を繰り返す。

- データ列 $\mathcal{D}_\mathcal{S}$ の中の $\{d_i, \cdots, d_{i+j(i)-1}\}$ を $\mathcal{E}_i$ とする。
- $\mathcal{E}_i$ の中に、ID が重複するデータが存在した場合それらを取り除く。
- $\mathcal{E}_i = \{e_1, \cdots, e_{|\mathcal{E}_i|}\}$ として、$1 \leq l < m \leq |\mathcal{E}_i|$ を満たす整数の組 l, m に対して

$$\begin{aligned} & M[n(e_1)][n(e_l)][n(e_m)] \\ \leftarrow\ & M[n(e_1)][n(e_l)][n(e_m)] + 1 \end{aligned} \tag{2}$$

とする。

この操作によって、$M[k][k'][k'']$ には、ID が k の ID を 1 周期分取り出したときに、ID が k' の ID が ID が k'' の ID より前に現れる回数が入力される。

次に $M[k]$ の各要素 $M[k][k'][k'']$ に対して、それが $\mathcal{D}_\mathcal{S}$ における s_k の出現回数と等しい場合 $M[k][k'][k''] = 1$ とし、そうでない場合 $M[k][k'][k''] = 0$ とする。すなわち

$$M[k][k'][k''] \leftarrow \begin{cases} 1 & M[k][k'][k''] = |\mathcal{D}_{\{s_k\}}| \\ 0 & (otherwise) \end{cases} \tag{3}$$

これにより $M[k]$ を、ID が k の ID を先頭としたときに各 ID が満たす順序関係の有向グラフとみなすことができる。すなわち、$M[k][k'][k''] = 1$ であるとき、ID が k の ID を先頭として次に ID が k の ID が現れるまでのデータ列の区分を抜き出すと、ID が k' のデータは必ず ID が k'' のデータよりも前に現れることを表している。

次に順序集合を取り出す上で無関係となる冗長な辺を取り除く。これは、例えば [2,3,5] という順序が成り立っているときに 2 が 3 の前に存在することと 3 が 5 の前に存在することが分かれば、2 が 5 の前に存在することは自明であるため、辺を省けることを意味する。このようにすることで、後で定常的な順序を満たす集合を取り出しやすくすることができる。

具体的には $M[k][k'][k''] = 1$ となる k, k', k'' について、$M[k][k'][k'''] = M[k][k'''][k''] = 1$ となる k''' が存在する場合

$$M[k][k'][k''] \leftarrow 0 \tag{4}$$

とする。これは、前述の条件を満たすときに、ID が k' の ID が ID が k'' の ID より前に来ていることは明らかなためである。最終的に、有向グラフ $M[k]$ に存在するパス上の点から構成される集合が $\mathcal{S}'$ となる。

以上のアルゴリズムを例とグラフの図を利用して説明する。表 3 は ID2 の ID を先頭に次に ID2 の ID が出現するまでのいくつかの列を表している。これは、上述したアルゴリズムの $\mathcal{E}_i$ に対応する。また ID の集合としては 1 から 5 までの整数を考えている。

表 3 特定 ID を先頭とするデータのサンプル

時系列区間 1	時系列区間 2	時系列区間 3
2	2	2
3	3	3
5	1	5
4	5	4
1	4	1
2	2	2

行列 M[2] は (3) の段階では、表 3 における 3 つの区間全てで k' が k'' より前に来ている場合に $M[2][k'][k''] = 1$ とするべきである。従って 1 となる成分は $M[2][2][1]$、$M[2][2][3]$、$M[2][2][4]$、$M[2][2][5]$、$M[2][3][1]$、$M[2][3][4]$、$M[2][3][5]$、$M[2][5][4]$ であり、この行列を $M[2]_{\text{original}}$ と書くことにすると

$$M[2]_{\text{original}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{5}$$

となる。図 1(a) は $M[2]_{\text{original}}$ を有向グラフを用いて表している。

続いて、冗長な辺について考える。前述の条件により、例えば $M[2][2][1] = 1$ は $M[2][2][3] = M[2][3][1] = 1$ であるために冗長な辺と言える。同様に、$M[2][2][4]$、$M[2][2][5]$、$M[2][3][4]$ が冗長な辺になっていて、これを $M[2]_{\text{redundant}}$ と書くことにすると

$$M[2]_{\text{redundant}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{6}$$

となる。これは、図 1(b) に有向グラフを用いて表されて

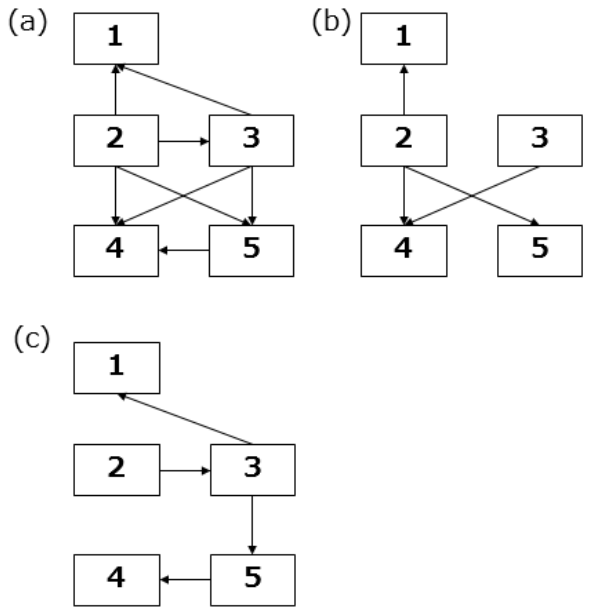

図 1 順序集合抽出の段階で定義される有向グラフ。(a)ID2 のメッセージを先頭とした集合を抜き出し、その集合での前後関係から定義した有向グラフ $M[2]_{\text{original}}$。(b) 有向グラフ $M[2]_{\text{original}}$ における冗長な辺の集合からなる有向グラフ $M[2]_{\text{redundant}}$。(c) 有向グラフ $M[2]_{\text{original}}$ から冗長な辺の集合である $M[2]_{\text{redundant}}$ を引いて、抽出すべき順序集合を表す有効グラフ $M[2]_{\text{sequence}}$。

いる。

最後に、冗長な辺を引くことで順序集合を現すグラフを構成することが可能になる。このグラフを $M[2]_{\text{sequence}}$ と書くことにすると

$$M[2]_{\text{sequence}} = M[2]_{\text{original}} - M[2]_{\text{redundant}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (7)$$

となる。これは有向グラフとして表現すると図 1(c) のグラフを得ることができる。このグラフは ID2 の点から辺を順番に辿ることで [2,3,5,4],[2,3,1] の順序集合を抽出することができる。実際に表 3 の全ての時系列データに対して、抽出した順序が崩れずに保持されていることがわかる。

実際には、ログを一度走査する過程で全ての ID を先頭としたグラフ $(M[1], M[2], M[3], M[4], M[5])$ を構成し順序集合が先頭の ID によらずに抽出された場合、それを正常な順序集合とみなす。例として、[2,3,5,4] が抽出された場合、[3,5,4,2],[5,4,2,3],[4,2,3,5] が抽出されるか検証し、抽出された場合にそれを最終的な結果とする。

4. 実車データを用いた手法の実験

この章では、実際の走行において取得した CAN のログに対して、本論文で提案した手法を適用した結果について述べる。

4.1 評価実験概要

CAN のログの取得は 3 つの異なる車種で行った。以下ではこれらを、車種 A,B,C とする。それぞれの車種に関して、合計 6 回の走行を行い、1 回につき数 100 万行程度のメッセージログを得た。走行を複数回行ったのは、各 ECU の起動のタイミングなどで順序が変わる可能性があり、そのような影響を受けずに正常な順序を保つ集合の抽出を狙ったためである。

4.2 評価実験結果

この節では、我々が実際に入手した CAN のログを用いて今までに説明したアルゴリズムを実証した結果を示す。代表的な車種では ID は 60 種類程度あり、それぞれの ID は固定された時間周期を持つ ID とそうでない ID に分類することが可能であった。時間周期は 10ms,20ms,40ms,100ms,200ms,300ms,500ms のものが存在し、時間周期で分類を行うと 3-10 程度の ID が一つのグループに属する結果が得られた。以下に 3 つの車種について、特定の時間周期を持つ ID の個数を示す。

表 4 特定の時間周期を持つ ID の個数

時間周期	車種 A	車種 B	車種 C
10ms	6	8	3
12ms	0	0	2
15ms	0	0	1
20ms	10	13	12
30ms	0	0	2
40ms	5	10	0
50ms	0	0	4
100ms	10	14	25
200ms	1	3	2
300ms	9	11	0
500ms	3	4	3
1000ms	0	1	6
周期あり合計	44	64	60
全 ID 数	44	64	67

車種 A,B に関しては全ての ID が一定の周期を持つことが確認されたものの、車種 C についてはいくつかの ID は一定の周期を持たなかった。周期の判定については、各 ID についてログにおける同じ ID のメッセージのタイムスタンプの値と着目しているメッセージのタイムスタンプの値の差分を取り、その分布に関するヒストグラムを作成することで行った。例えば 40ms の周期を持つ ID に関しては図 2(a) のように一つの値に大きなピークを持ち、数 ms 程度の分散しか持たないという性質を持つ。一方で、特定の周期を持たない ID に関しては、タイムスタンプの差分

の分布が図 2(b) のように複数のピークを持つ場合などがあった。

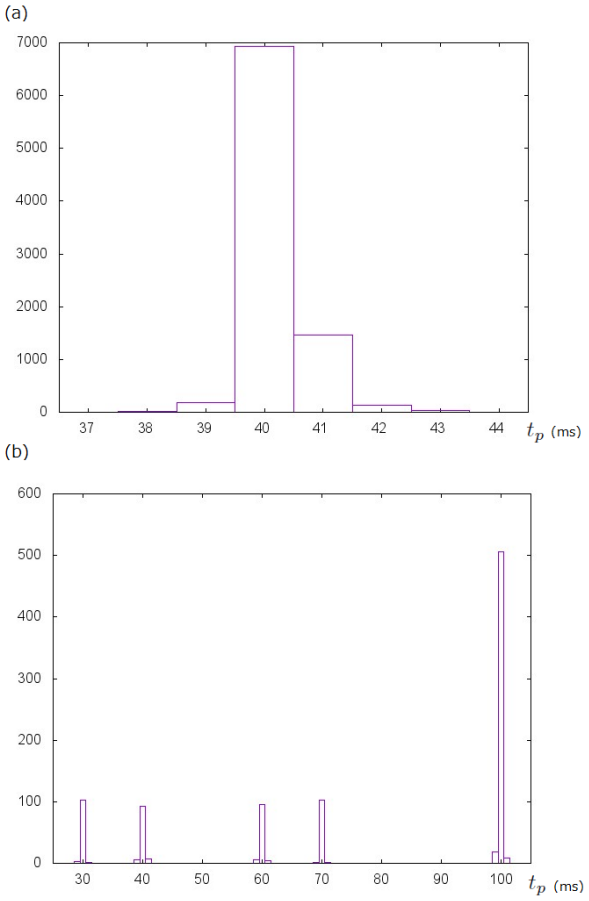

図 2 (a)40ms の時間周期を持つ 1 つの ID のメッセージに注目し、前のメッセージとの時間間隔をヒストグラムにしたもの。(b) 特定の時間周期を持たない 1 つの ID のメッセージに注目し、前のメッセージとの時間間隔をヒストグラムにしたもの。

また、一定順序集合として抽出された集合に含まれた ID の個数を表 5 に示す。一つの欄に複数数字があるものは、一つの時間周期に対して複数の集合が抽出されたことを意味する。例えば、アルゴリズムの説明で例としてしめしたように [2,3,5,4],[2,3,1] が抽出された場合、この集合の大きさがそれぞれ 4 と 3 であるため、4,3 と表記する。この例の ID2 と ID3 のように一つの ID が複数の集合に属することもあるため、順序集合の要素数の合計（この例では $4+3=7$) が時間周期で分類した集合の要素数より大きくなる場合もある。

各時間周期の ID の集合に対して、一定の順序を満たす集合が存在する場合としない場合があり、存在する場合は平均的には 4 つ程度のメッセージが常に変わらない順序を満たしながら送信されている様子が明らかになった。CAN においてこれらは、関連性のある ECU が送信するメッセージ群と考えられる。

表 5 抽出された順序集合のサイズ

時間間隔	車種 A	車種 B	車種 C
10ms	4	4	-
12ms	-	-	-
15ms	-	-	-
20ms	5,3,3	6	4,4,4
30ms	-	-	-
40ms	-	-	-
50ms	-	-	-
100ms	4,4,4,4,4	-	10,4,6
200ms	-	-	-
300ms	3,5,5,6,6	8	-
500ms	-	-	-
1000ms	-	-	-

5. 異常検知手法の提案

本節では、今までに挙げた CAN のログの性質を利用した異常検知の手法を提案する。正常状態において一定の順序を満たす集合をログから抽出できたら、異常検知は各集合における各 ID の順番をデータベース化することで可能になる。ここで、順番とは各集合において該当する ID が何番目に現れるかという指標であって、集合 [2,3,5,4] における ID3 の順番は 2 である。異常の検知は以下のプロセスで行う。

(1) 各集合ごとにカウンタを設定し、初期値を -1 とする。以降集合に属する ID を受信した際に (2) 以下の処理を行う。

(2) その集合に対応するカウンタが -1 の場合、受信 ID の順番をカウンタに代入。

(3) カウンタが -1 で無い場合、カウンタの値を c、ID の集合内での順番を i、集合のサイズを S とした時に

$$c+1 \equiv i \mod S \tag{8}$$

であるならば正常と判定しカウンタの値を i で置き換える。上の等式が成り立たない場合異常と判定し、異常時の処理を行う。

この手法を用いると、例えば ID3 のメッセージ後に ID5 のを受信した場合、式 (8) の左辺と右辺は 3 となり合同式が成り立つ。ID4 のメッセージの後に ID2 のメッセージを受信した場合も左辺と右辺が 5 と 1 となって、合同式が成り立つ。一方で、ID3 の後に ID4 を受信した場合、左辺と右辺は 3 と 4 になるため合同式は成り立たず、異常として検出される。ここで、間にこの集合とは関係のない ID が入っている場合は上で説明したプロセスを行わないことに注意する。したがってカウンタの更新も行われない。これは、実際の CAN のデータにおいても正常動作時に、順序集合の間に別の ID が入り込むことは頻繁に起こるためである。

実際にログの順序を意図的に正常な順序から変更したログに対してこの手法を用いて異常検知を行った結果、順序

を変更した行を検知することを確認している。

6. まとめと今後の展望

本論文ではタイムスタンプとメッセージ ID を持つログについて、一定の順序を満たしている ID の集合を抽出する手法を提案した。提案したアルゴリズムは、ログを 1 度操作し前後関係に基づいた有向グラフを構成するため、ある時間周期を持つ ID の集合のサイズの多項式時間で計算を行うことができる。この手法を用いて CAN のログに対して順序集合の抽出を行ったところ、複数の ID が一定の順序を満たす集合として抽出することが確認された。

本論文で抽出した順序集合は、ある ID が別の ID より「必ず」前に出現するという条件を満たしている。例えばこれを「90 %以上」前に出現するという条件に変更することが可能で、この場合抽出と検知のアルゴリズムを若干変更する必要がある。実際、システムにおいて正常な状態において、必ず一定の順序を満たす集合が存在することは稀で、確率的な要素を含めた方が適用範囲は広いと考えられる。CAN のログにおいても、高い確率で保たれる順序集合の存在は見られたものの、本論文で具体的には取り扱わなかった。抽出部分のアルゴリズムは 3 式において 1 とする条件を緩和することで実現できると考えられる。また、この場合検知部分のアルゴリズムはメッセージ一つ一つに対して検知を行うのではなく、複数のまとまったメッセージ群に対し、特定の順序が保たれる割合を計算し、それを元に異常検知を行う手法が必要になる。このことは、確率を導入した場合に抽出結果がどのように変わるかなどを含めて今後の研究課題とする。またここでは、異常時にどのような処理を行うべきかという点については本論文の範囲外とし今後の課題とした。

参考文献

[1] T. Hoppe, S. Kiltz, and J. Dittmann: Security threats to automotive CAN networks—Practical examples and selected short-term countermeasures, In Proceedings of the 27th international conference on Computer Safety, Reliability, and Security, SAFECOMP '08, pp. 235-248, (2009).
[2] K. Koscher, A. Czeskis, F. Roesner, S. Patel, and T. Kohno: Experimental security analysis of a modern automobile, IEEE Symposium on Security and Privacy 2010, pp. 447-462, (2010).
[3] S. Checkoway, D. McCoy, B. Kantor, K. Koscher, A. Czeskis, F. Roesner, and T. Kohno: Comprehensive experimental analyses of automotive attack surfaces, the 20th USENIX Security Symposium, (2011).
[4] C. Miller, C. Valasek. Adventures in Automotive Networks and Control Units DefCon 2013
[5] C. Miller, C. Valasek. Remote Exploitation of an Unaltered Passenger Vehicle Defcon 2015
[6] D. K. Nilsson, U. E. Larson, and E. Jonsson, Efficient In-Vehicle Delayed Data Authentication Based on Compound Message Authentication Codes. Vehicular Technology Conference, 2008.
[7] O. Hartkopp, C. Reuber, and R. Schilling. MaCAN - Message Authenticated CAN Embedded Security in Cars, 2012.
[8] 氏家良浩, 岸川剛, 芳賀智之, 松島秀樹, 田邉正人, 北村嘉彦, 安齋潤車載ネットワークにおける CAN フィルタの提案. Symposium on Cryptography and Information Security, SCIS 2015.
[9] 桑原拓也, 馬場雪乃, 鹿島久嗣, 氏家良浩, 岸川剛, 鶴見淳一, 芳賀智之, 松島秀樹 CAN メッセージ頻度に注目した車載ネットワークの統計的異常検知. Symposium on Cryptography and Information Security, SCIS 2016.
[10] 芳賀智之, 氏家良浩, 鶴見淳一, 岸川剛, 前田学, 松島秀樹, 安齋潤クラウドを利用した車載ネットワーク向け統計的異常検知システムの提案 Symposium on Cryptography and Information Security, SCIS 2016.
[11] M. Muter and N. Asaj: Entropy-Based Anomaly Detection for In-Vehicle Networks. IEEE Intelligent Vehicle Symposium. (2011).
[12] S. Ahn, H. Kim, J. Jeong, K. Kim. A Countermeasure against Spoofing and DoS Attacks based on Message Sequence and Temporary ID in CAN. Symposium on Cryptography and Information Security, SCIS 2016.