深層学習を用いたインターネットトラフィックの異常検知手法 (1/2)

和久井拓, 近藤賢郎, 寺岡文男　　{dona, latte, tera}@inl.ics.keio.ac.jp

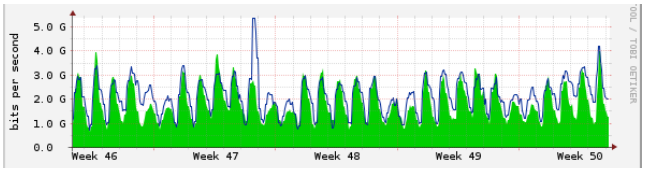
● インターネットバックボーンにおける異常検知

- インターネットバックボーンの特徴
  - 様々なネットワークを相互接続する大規模なネットワーク
  - 多種多様なサービス・ユーザのトラフィックが大量に流れる
    →平常時でも時間軸に対する変動が大きい
  - 様々な要因の異常事象が発生 (e.g., 機器故障, サイバー攻撃, イベント)
- 異常検知機構の必要性
  - 機器故障やイベントが要因の場合: ユーザやサービスに影響の恐れ
  - サイバー攻撃が要因の場合:ユーザやサービスに影響 / 加害者になる恐れ

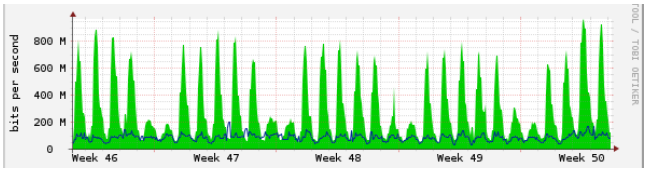
⇒異常事象を管理者に通知し，早期対処が必要

● インターネット・トラフィックの特性

- 局所性 (locality)
  - 宛先アドレスが観測点から近い程トラフィック流量が多い
  - ユーザが利用するサービスの局所性に起因
- 周期性 (periodicity)
  - トラフィック流量の周期性
  - 曜日ごとに異なる特徴



国内向けトラフィック



国際向けトラフィック

● 深層学習による異常検知

- トラフィックの振るまいを学習・予測
  - 平常時のトラフィックの特徴 (振るまい) を学習
  - 学習結果に基づいた予測と実トラフィックを比較して異常検知
- 未知の脅威にも対応可能
- スケーラビリティの問題
  - インターネットで観測される通信フロー毎の学習・予測は非現実的
  - (異常検知性能を損なわない) トラフィックの集約方法が必要

深層学習を用いたインターネットトラフィックの異常検知手法 (2/2)

和久井拓, 近藤賢郎, 寺岡文男　　{dona, latte, tera}@inl.ics.keio.ac.jp

提案手法 GAMPAL: General purpose Anomaly detection Mechanism using Path Aggregate without Labeled data

- 経路情報 (AS_PATH) によりトラフィックを分類
  - インターネット・トラフィックの局所性に基づいた集約方式
- LSTM (Long Short-Term Memory) -RNN によるトラフィック予測
  - インターネット・トラフィックの周期性を反映
- 予測トラフィックと実測トラフィックの振る舞いを比較
  - トラフィック流量に基づく比較 (多種多様な異常事象を汎用的に検知)
  - 収集が困難なラベル付きデータが不要

経路情報 (AS_PATH) による集約方式

- AS_PATH: 宛先 IP アドレスによって定まるトラフィックの経路
- 観測点の視点で AS_PATH が 3 つ目まで共通するものを集約

観測されるトラフィック情報

Time	src IP	dst IP	bytes
00:00	x.x.x.x	1.0.0.1	6
00:02	x.x.x.x	1.0.4.1	18
00:02	x.x.x.x	1.0.5.1	5
00:04	x.x.x.x	1.0.0.4	6
00:06	x.x.x.x	1.0.0.1	15
00:12	x.x.x.x	1.0.0.4	3
00:14	x.x.x.x	1.0.6.1	6

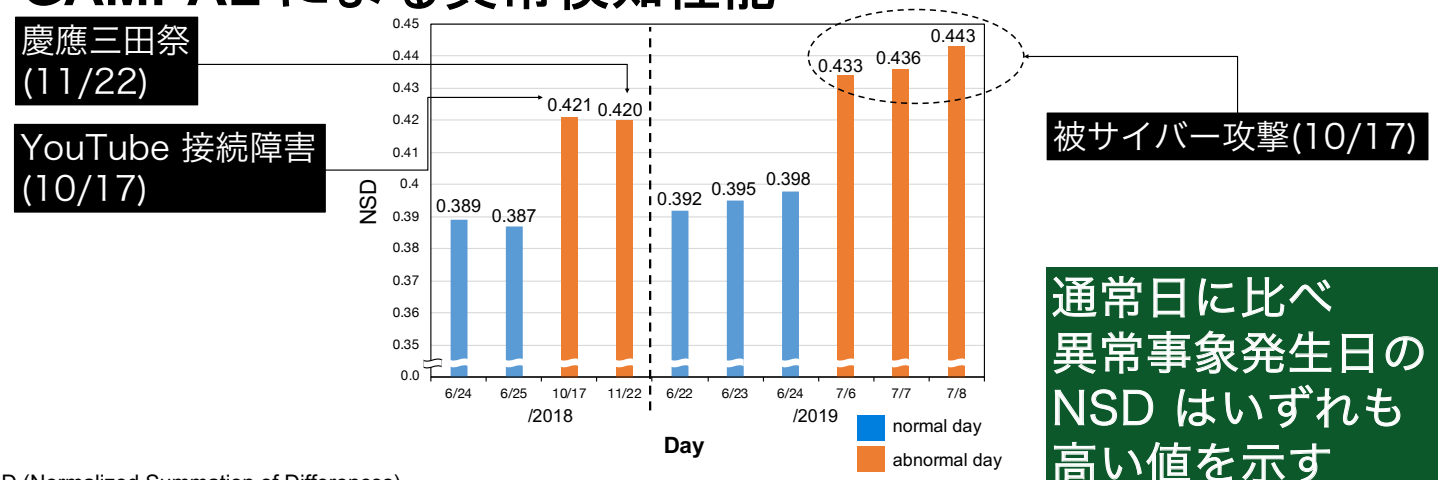
AS_PATH でクラス分けされた Prefix

インデックス	AS_PATHの観測点側から 3 つ目まで	Prefix
1	4713 2914 13335	1.0.0.0/24
2	4713 2914 15412	1.0.4.0/24 1.0.6.0/24
3	2497 2519	1.0.5.0/24

終点 IP でクラス分けされたトラフィック情報

時間	始点IP	終点 IP	バイト数
00:00	x.x.x.x	1.0.0.1	6
00:02	x.x.x.x	1.0.4.1	18
00:02	x.x.x.x	1.0.5.1	5
00:04	x.x.x.x	1.0.0.4	6
00:06	x.x.x.x	1.0.0.1	15
00:12	x.x.x.x	1.0.0.4	3
00:14	x.x.x.x	1.0.6.1	6

GAMPAL による異常検知性能



通常日に比べ異常事象発生日の NSD はいずれも高い値を示す

＊NSD (Normalized Summation of Differences)
実測/予測トラフィックの差異を評価する独自指標

通常日と異常事象発生日の NSD＊の平均