分野を超えた視点で異分野エンジニアが NWの異常検知に挑戦

～高精度ネットワーク監視への工夫と課題～

ソフトバンク株式会社

内海、合志、平野

2024/07/03

SoftBank

自己紹介

内海 友輔

出身
- 岩手県一関市

趣味
- フルマラソン（サブ3.5くらい）

担当業務
- Telemetry収集, 検証
- データ分析用の基盤構築/運用
  （オンプレ/パブリッククラウド）

経歴
- 2018~2022:
  - プライベートクラウド構築/運用
  - IaC、CICD環境構築/運用
- 2022/4~: 現業務

合志 和真

出身
- 熊本県合志市

趣味
- コーディング、読書

担当業務
- Telemetryコレクタの開発
- トラフィックの傾向異常検知導入

経歴
- 2020~2023:
  基地局監視システムの保守・運用
- 2023/4~:
  現業務

平野 泰平

出身
- 宮城県仙台市

趣味
- テニス、スキー（テクニカル 目標）

担当業務
- 異常検知機能開発
- 作業自動化支援機能開発

経歴
- 2001~2020:
  SIerにてサーバインフラ開発
- 2020/11~:
  現職

我々の組織の立ち位置の紹介

SoftBank

依頼, 相談

NW機器
導入主管部

NW機器
運用主管部

所属組織

実装, 提案

NW機器の検証導入や保守運用

よろず屋、何でも屋

- NW機器の異常検知、元となるデータ収集、トリガーとなる検知の立て付け、‥
- 様々なデータソースで何ができるか、NW機器の担当と話したり

→ 我々、実はNW経験あまり無い人たちです。お手柔らかに。。

アジェンダ

1. 取組みの背景

2. NW障害の早期検知
   a. リアルタイム検知　（内海）
   b. 独自コレクタ開発　（合志）
   c. 傾向異常検知　（平野）

3. まとめと議論のポイント

取組みの背景

生成AIが作成してくれた通信インフラの重要性を表すイメージ図

通信インフラの重要性を表すイメージを数枚作成し説明文をつけてください

◆緊急サービス

緊急事態にスマートフォン、
インターネット通信ツールを使用して
危機へ連携して対応している様子

◆日常生活への影響

自宅で働く人、SNSを利用する人、
オンラインショッピングをする人、
緊急電話をかける人のシーン

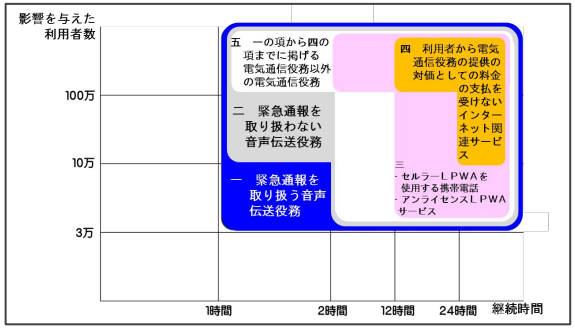
◆経済活動への影響

リモートで行われるビジネス会議、
クラウドコンピューティング、
オンラインでの国際取引のシーン

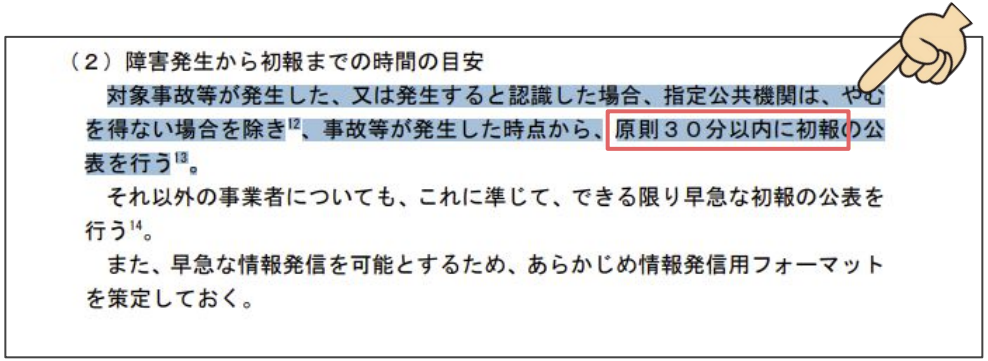
ネットワークインフラを取り巻く環境の変化

- NWサービスの重要性がますます顕著に。
  NW障害発生から初報まで、**30分以内**というガイドラインも制定される。
- 障害復旧はもちろん、HP掲載までの短時間化にも努めている。
  → 障害の**知得・復旧に期待される時間が日に日に短く**なっている。

良いことではある



（総務省）
電気通信政策の推進 > 安全・信頼性の向上 > 重大な事故の報告



（２）障害発生から初報までの時間の目安

対象事故等が発生した、又は発生すると認識した場合、指定公共機関は、やむを得ない場合を除き[12]、事故等が発生した時点から、原則３０分以内に初報の公表を行う[13]。

それ以外の事業者についても、これに準じて、できる限り早急な初報の公表を行う[14]。

また、早急な情報発信を可能とするため、あらかじめ情報発信用フォーマットを策定しておく。

（総務省）
電気通信サービスにおける
障害発生時の周知・広報に関するガイドライン（令和5年3月）

[1]: https://www.soumu.go.jp/menu_seisaku/ictseisaku/net_anzen/jiko/judai.html
[2]: https://www.soumu.go.jp/main_content/000869357.pdf

復旧自動化を取り巻く環境の変化

総務省の電気通信事故に関する文言の変化

2023年2月	障害事象の検出、被疑箇所の特定及び復旧措置のそれぞれの**迅速化・自動化を推進**することで事故の長期化を防止する対策を徹底すること。[1]
2022年12月	復旧手順の明確化・**自動化**を行い、事故の長期化の防止のための対策を徹底すること。[2]

自動化に対する否定的な考えはほぼ消滅

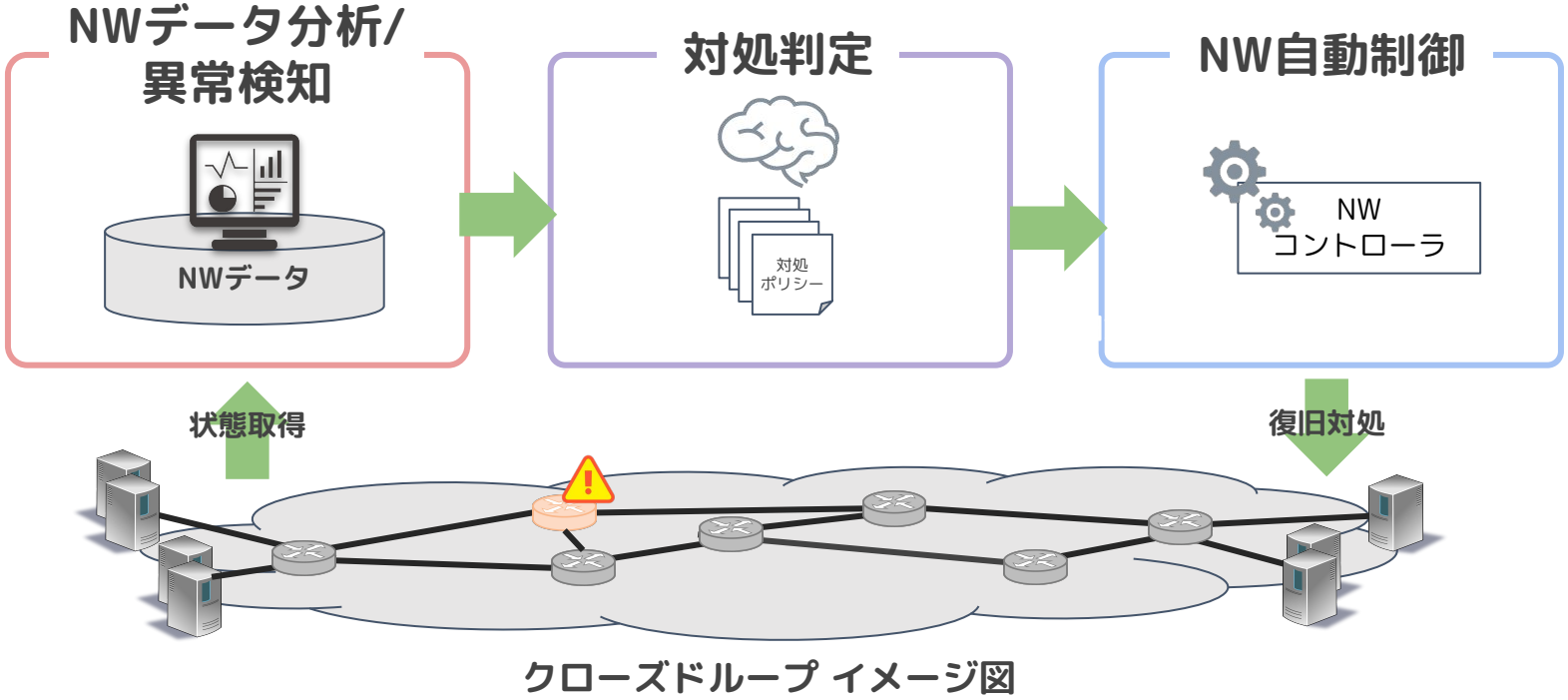
と、思っています

- むしろ、「何故やらないのか」という風潮にまでなっている。

[3]: https://www.soumu.go.jp/main_content/000860727.pdf
[4]: https://www.soumu.go.jp/main_content/000849780.pdf

早期復旧を実現する要素とは‥？

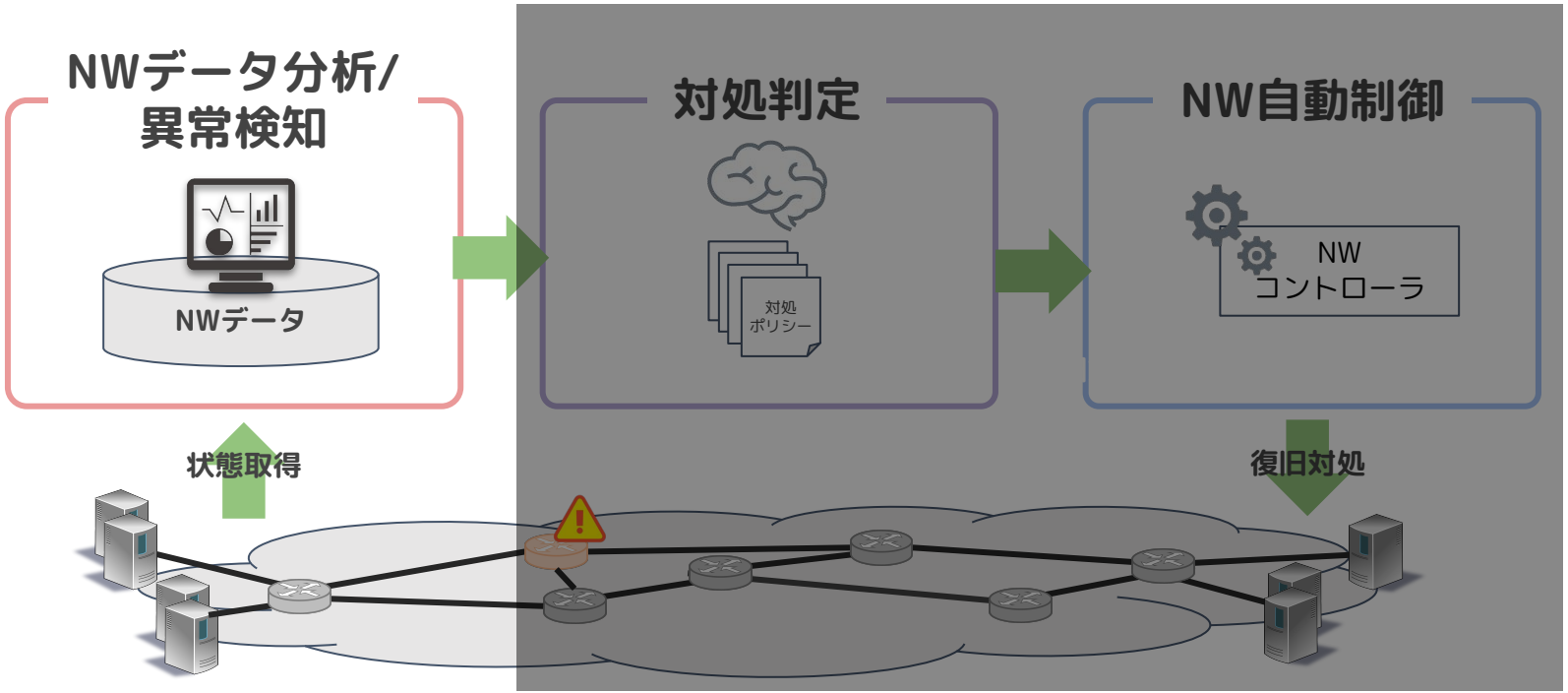
NW障害検知～自動復旧のクローズドループの実現が必要



クローズドループ イメージ図

早期復旧を実現する要素とは‥？

今回発表のスコープ



クローズドループ イメージ図

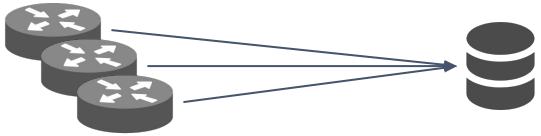
リアルタイム検知

リアルタイム検知に必要な事は

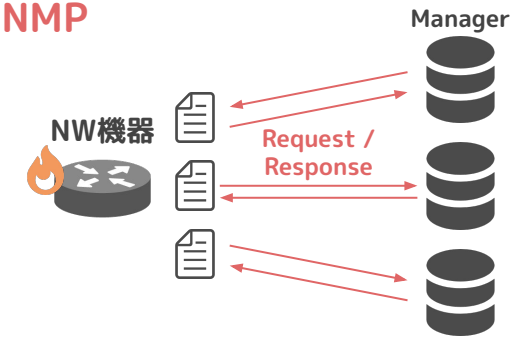
SoftBank

“1分・1秒でも早く”
検知の早期化を考える必要性

高頻度 & 効率的なデータ取得

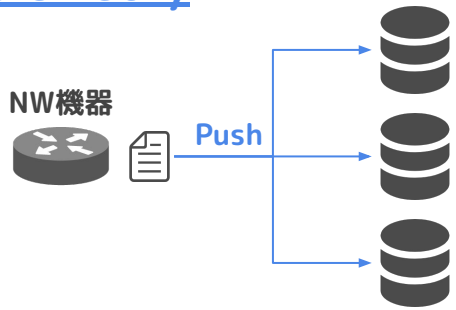


SNMP



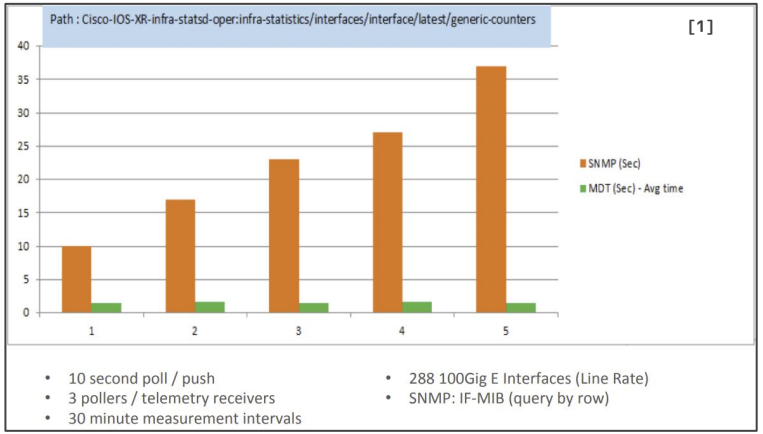
現実的にはおよそ5分間隔

Telemetry



より短期間での収集（<< 5分）

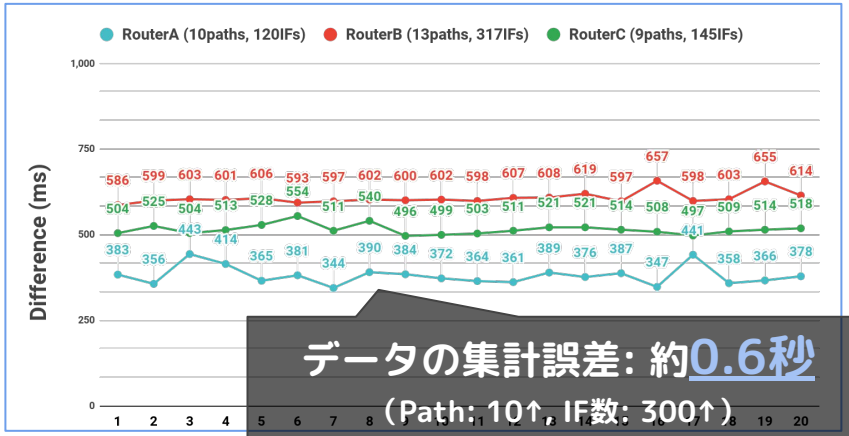
（おさらい）Telemetryとは



特徴（期待されること）

- Pub-Sub形式: 効率的なデータ取得
- データ取得の高頻度化
- 多種多様な取得可能データ
  他にも‥

複数データ取得時の集計時間差



多数のデータを取得していても、
集計にほぼ時間差がない

サイレント障害の検知に有用

[5]: https://archive.nanog.org/sites/default/files/Cadora_Ten_Lessons.pdf

とはいえ、Telemetryどれぐらい使ってますか？？

SoftBank

なんで導入されないのか？

SNMPでも
良いんじゃないか？

現状より良くなる
イメージが無い

データ量増加に伴う
新たな課題
（インフラ面など）

マルチベンダ環境では
センサパスが共通ではない

過去のJANOGにおけるTelemetryの振返り

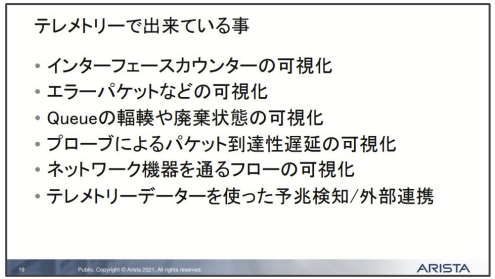
SoftBank

Telemetry Working Group 最終報告（JANOG 43）

・サンプリング間隔毎の
バーストトラフィック検出

・NWの視点だけではなく、
インフラ全体の考慮が必要

In-band Telemetry -データプレーン遅延時間観測（JANOG49）



Telemetryのユースケース例

・IFカウンタの可視化
・エラーパケット等の可視化

gNMIcを活用したマルチベンダー環境でのテレメトリ技術の実践（JANOG 52）

Q: リアルタイム性は本当に欲しい？

- 我々としては欲しい。
- 対処や人間の判断を含め、フロー全体で考えると検知にはあまり猶予が無いと思っている。

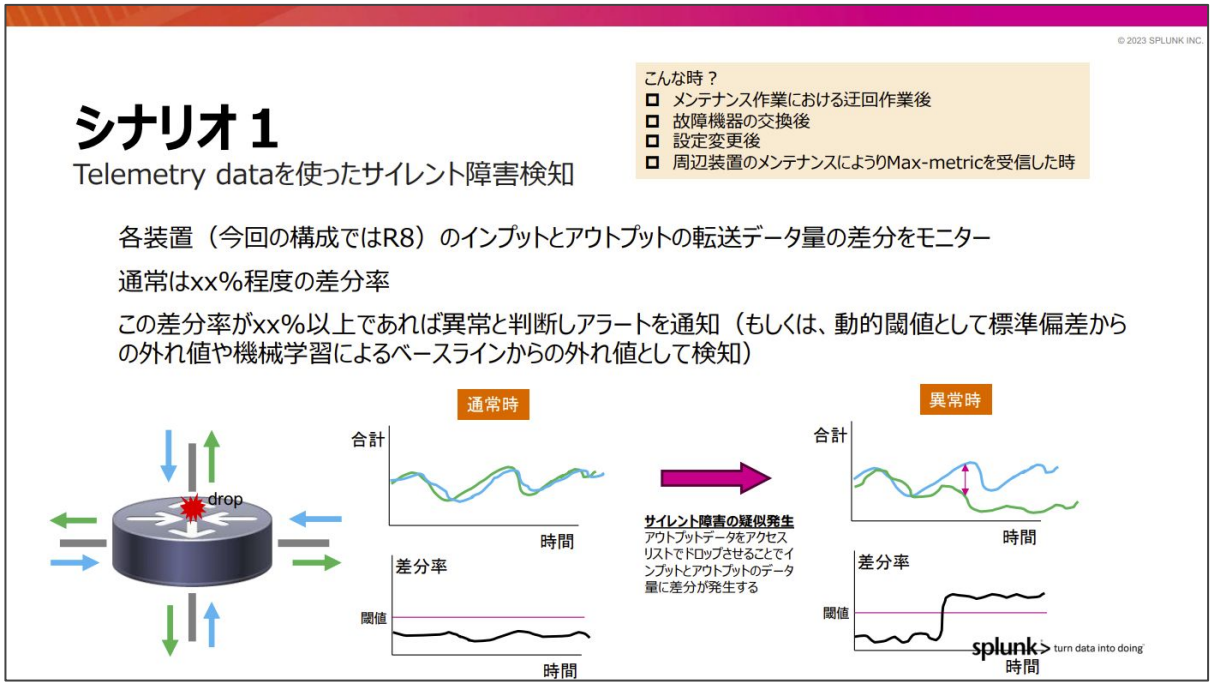
Telemetryを商用監視に適用して、実際に直面した課題についてお話したい

[6]: https://www.janog.gr.jp/meeting/janog43/program/telewg/
[7]: https://www.janog.gr.jp/meeting/janog49/intele/
[8]: https://www.janog.gr.jp/meeting/janog52/gnmic/

Telemetryデータを用いたサイレント障害の知得（JANOG 53）

SoftBank

- 前回のJANOGでも**サイレント障害の知得**に関する議論がなされていた。
- 今回は現場レベルでの取組みについて、この場で共有したい。

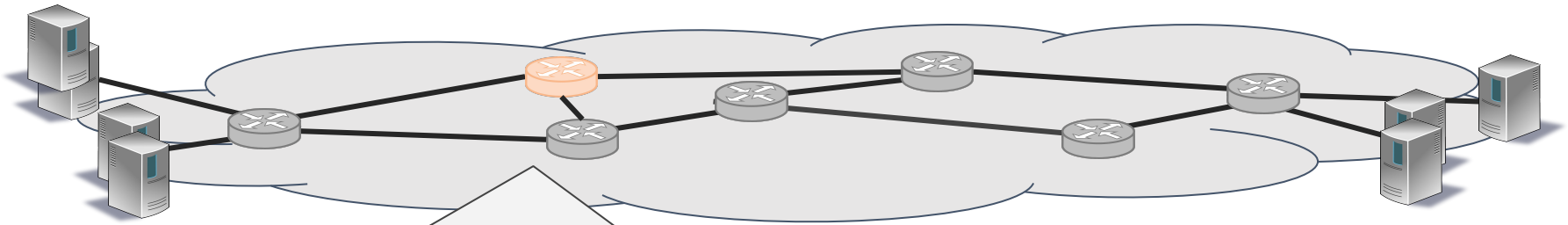


[9]: https://www.janog.gr.jp/meeting/janog53/nwviz/

Telemetryでサイレント障害知得をするモチベーション

SoftBank

例）中継区間のNW機器における<u>ルーティングテーブルの欠落</u>



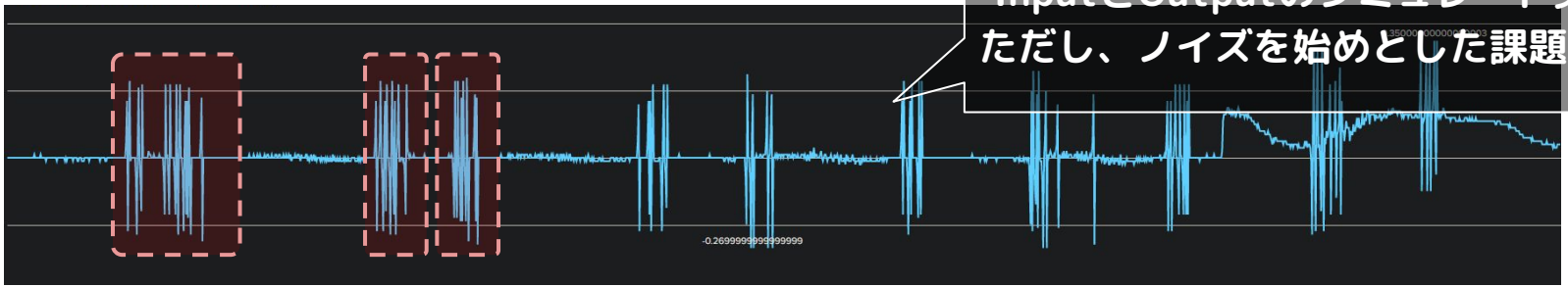
<u>この障害想定パターンに限って</u>は、
ルーティングテーブルの経路数を監視する等‥

じゃあxxは？

yyは考えた？

<u>パターン網羅は非現実的</u>
かつ <u>後追いでの監視実装</u>

▶ そこで、SNMPでのサイレント障害の知得を試みたものの。

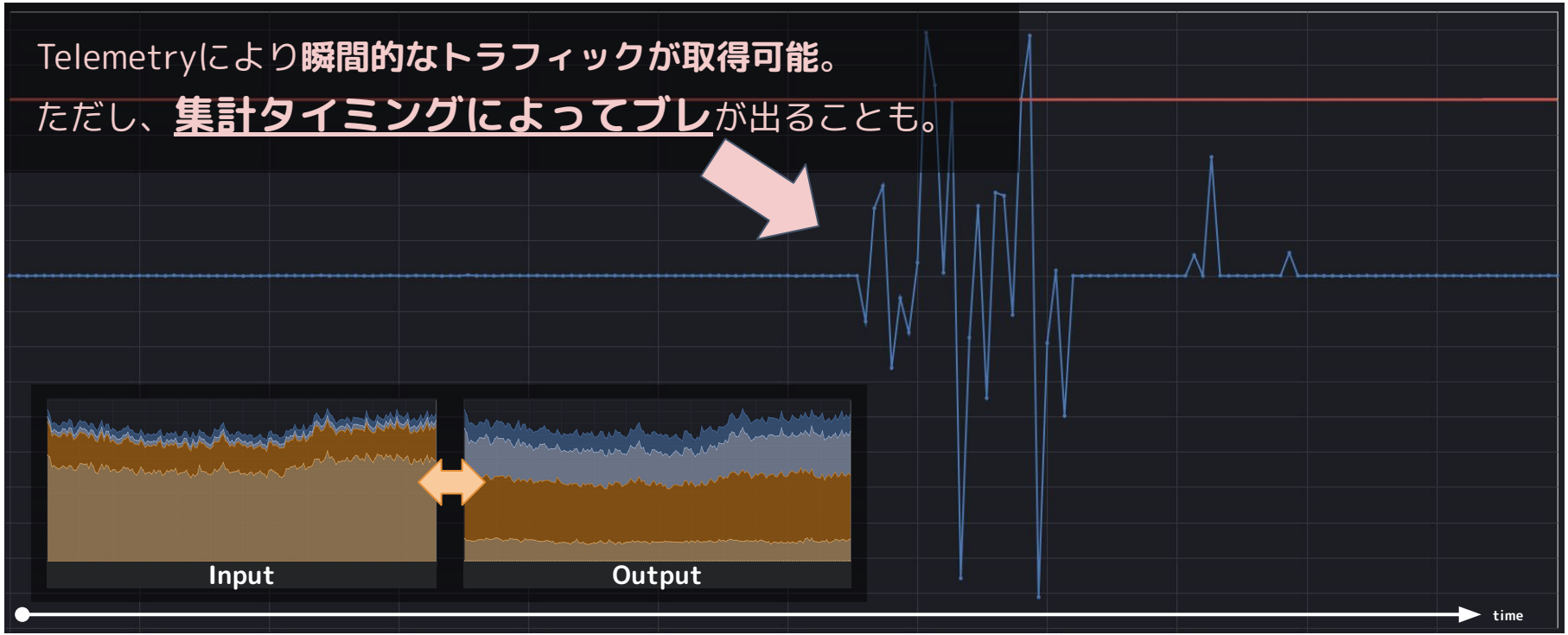


- InputとOutputのシミュレート実施。
ただし、ノイズを始めとした課題に直面

サイレント障害の知得に向けた精度向上の取組み

SoftBank

ただし、Telemetryをそのまま導入するだけでは駄目だった。



サイレント故障検知について現状抱えている課題

SoftBank

異常と判断する基準、閾値の定め方

過去実績から定める（静的）	機械学習で定める（動的）	…

※ NWが常に安定しているわけではない。
実際は作業等で**NW傾向が随時変動**する。

他のTelemetry利活用事例（劣化具合での検知）

- 長いスパンでの収集では**Errorの値が平滑化され**、検知が遅くなってしまう

- Errorを細かい粒度で取得できる（解像度が上がった）ため、対処が短時間化

※ 技術的、NW機器のリソース負荷、両面でメリット

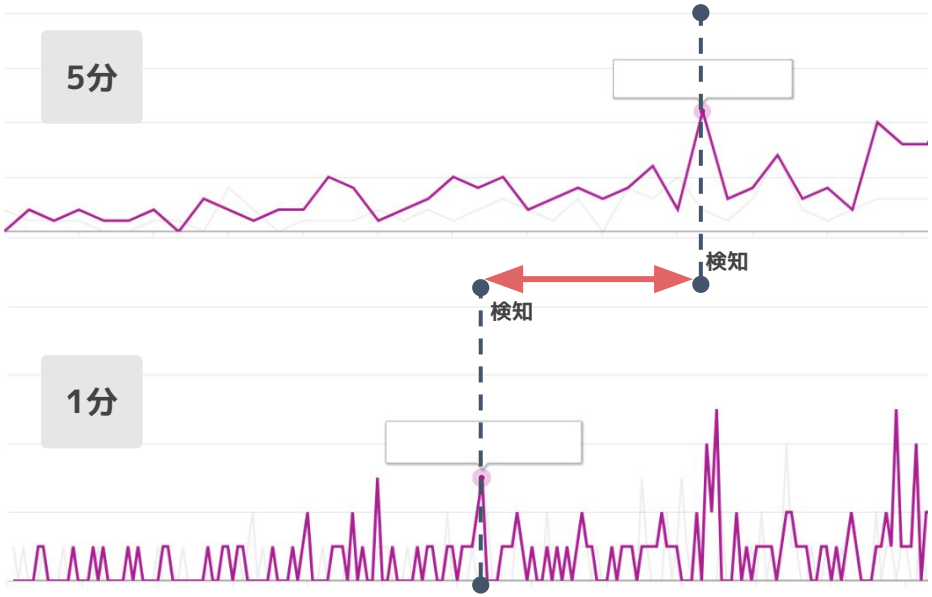
あるIFの劣化具合における検知比較

SoftBank

同等ロジックにおける、集計期間での検知時間の差を比較

- 5分集計データ
- 1分集計データ

特定IFにおける劣化具合



- 5分集計パターンだと検知が遅い？
- 1分集計パターン程の早期検知は不要？

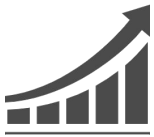
この判断基準の制定は難しい。

予兆の検知？ or 障害の検知？

Telemetryデータ取得/利活用以外における課題

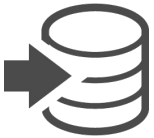
SoftBank

データ量の肥大化



- サーバリソース
  + ストレージ料金
  + データ転送の帯域..
- 不要なデータは
  前段で削除したい。
  とはいえ、後から
  必要になることも。。

データの保持場所

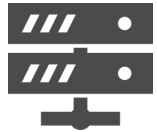


- オンプレミス
  → 保守運用コスト
- クラウド（IaaS）, SaaS
  → データ転送量に
  対する従量制課金
  → ストレージ料金

データ取込み処理

- 事前のETL処理が必要
- **同じイベント内**での計算
  例: %の算出
- **異なるイベント間**の計算
  例: カウンタ値から
  差分算出 → pps

インフラ負荷考慮



- 検知のため
  全IFを集計している。
- なおかつ、高頻度で
  サーチすることによる
  インフラの高負荷

独自コレクタ開発

コレクタを必要とする背景

SoftBank

早期検知・復旧が組織の必達目標

- 何としても障害を**早期**（1秒,1分でも早く）に検知したい
  - 現状のSNMPで取れるログ＋**障害の前兆**を把握するためのトラフィック等の情報が必要
- **傾向異常検知**では複数点の異常トラフィックのデータが必要
  - SNMP（5分毎）だと異常データ3つ取るだけで15分かかってしまう
  - Telemetry（1分以下）だと長くても3分で取得可能

今後、早期検知・復旧の温度感は上がっていく

- 社内の全てのルータ（n万台）から上記の情報を取得したい
  - 多種多様な多数のルータの一元管理

既存のコレクタは？

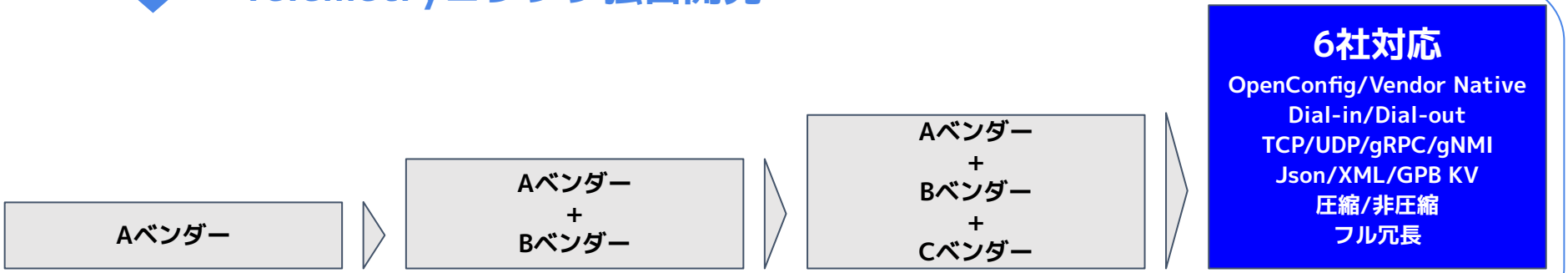
SoftBank

- Pipeline
- Telegraf
- gNMIc

最初はOSSのコレクタを使用
微妙に帯に短かし たすきに長しだった
※今は違うかもしれないし、我々の使い方が悪かったかもしれない

→多少コストがかかっても
自由に運用できるコレクタを開発した方が今後活きるのでは？

Telemetryコレクタ独自開発へ



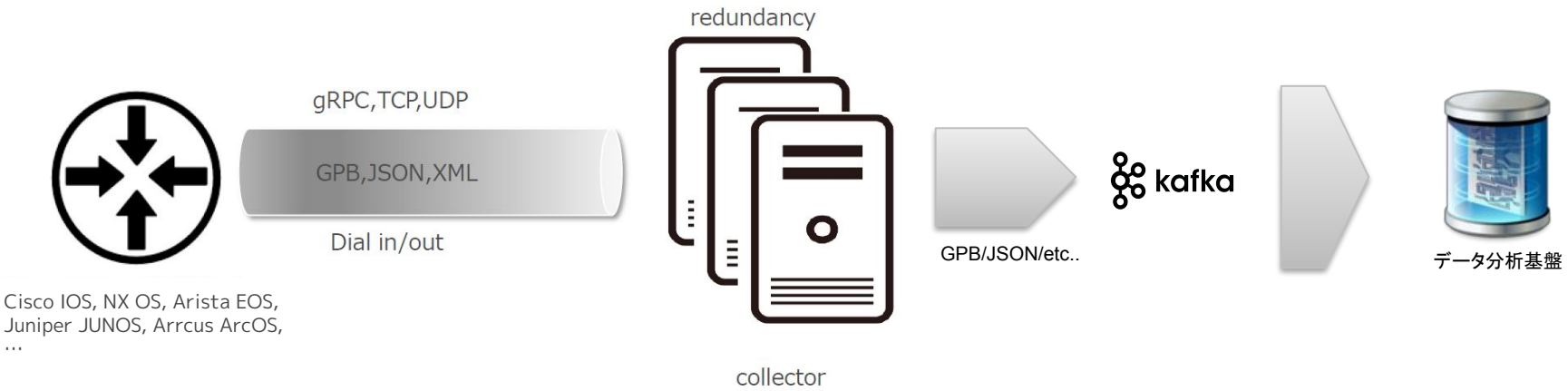
一つ一つ対応していき独自のコレクタを作成

SoftBank

コンセプトは Telemetryデータ取得の一元化

- **多様なルータ** から、**多様なデータ** を、**リアルタイム** で取得したい
  - **社内導入済の各ベンダ** から、**Telemetryデータ** を、**高頻度** でデータ取得

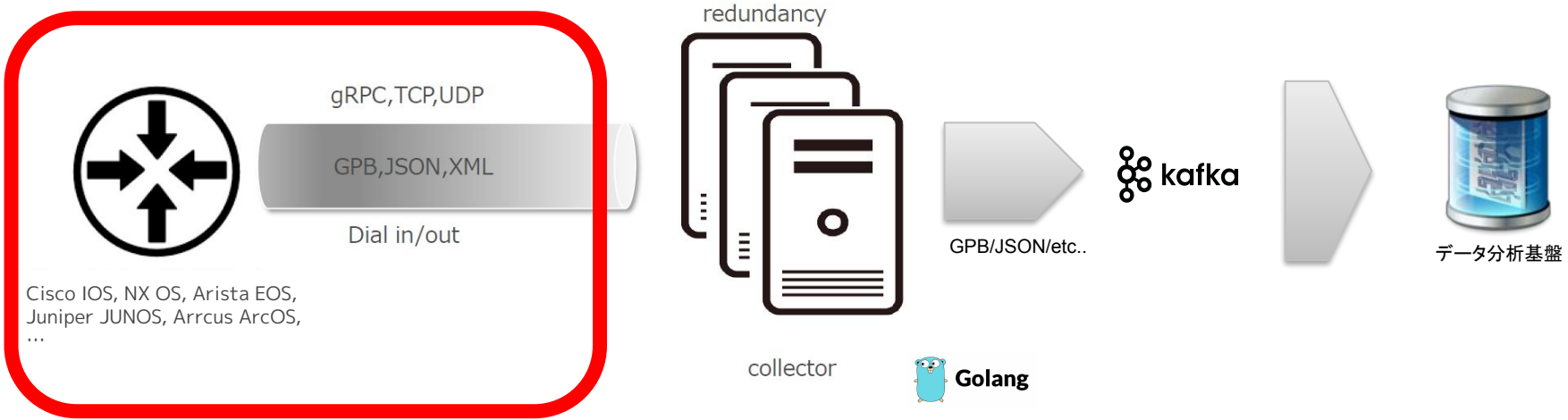
- 取得したデータは遅滞なく確実に後段へ転送



SoftBank

複数ルータ対応

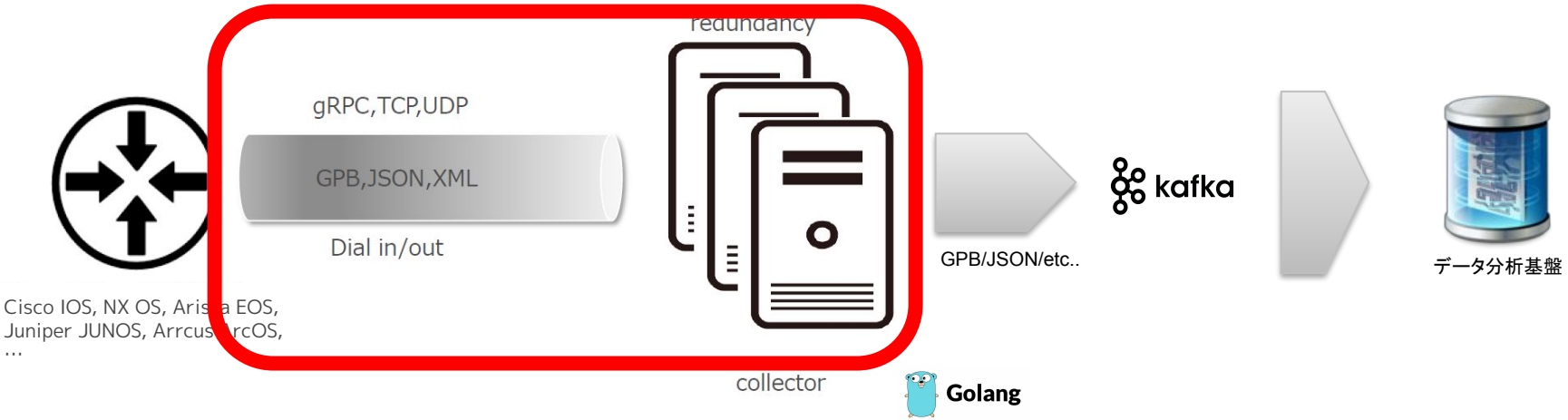
- Cisco gRPC, Juniper gRPC(認証方式が違う), gNMI, …
- 取得方式（Dial-In / Dial-Out）, データ構造（GPB / JSON, …）の差を吸収
  - プロファイルごとにロジックを分け、制御できる仕組み



SoftBank

耐障害性（ルータ側 down / S.O. 後のデータ取得自動再開

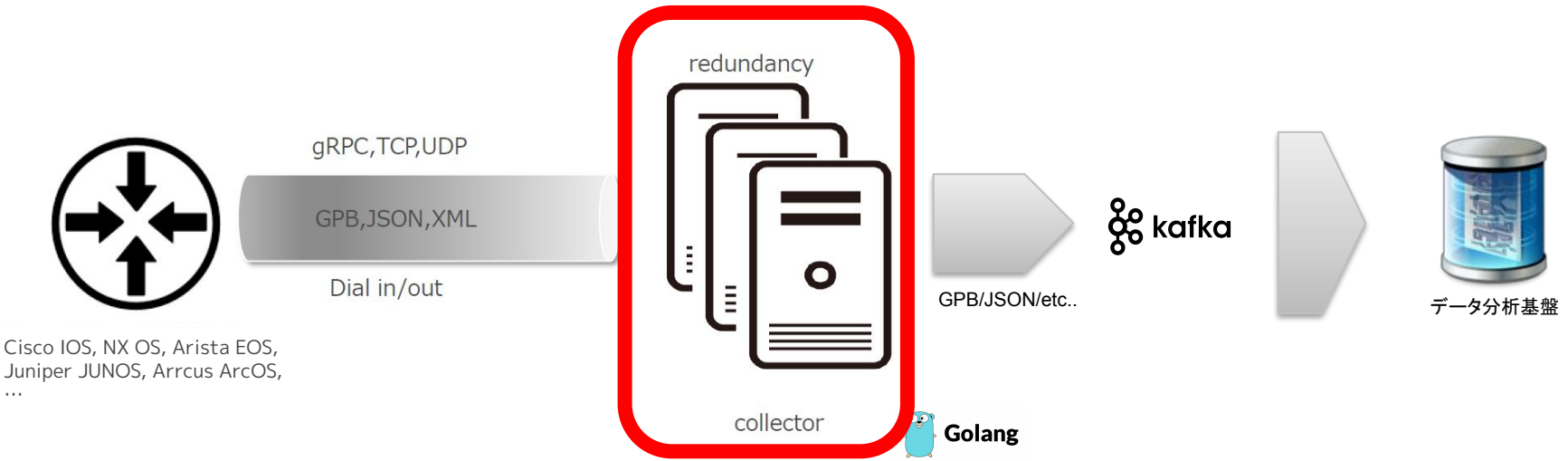
- ベンダ、ルータ毎に復旧までの時間が違うので、その辺りも柔軟にしたい
- 5桁オーダの数のルータをコレクタ内で上手く分散処理させたい
  - コレクタ追加、障害時にも自動でロードバランシングする仕組み



SoftBank

様々な用途のデータを一様に管理

- タグ付け機能などでデータを管理
- ルータごとの収集・転送設定をプロファイルで管理
  - タグをつけたデータをkafkaの特定Topicに転送、など

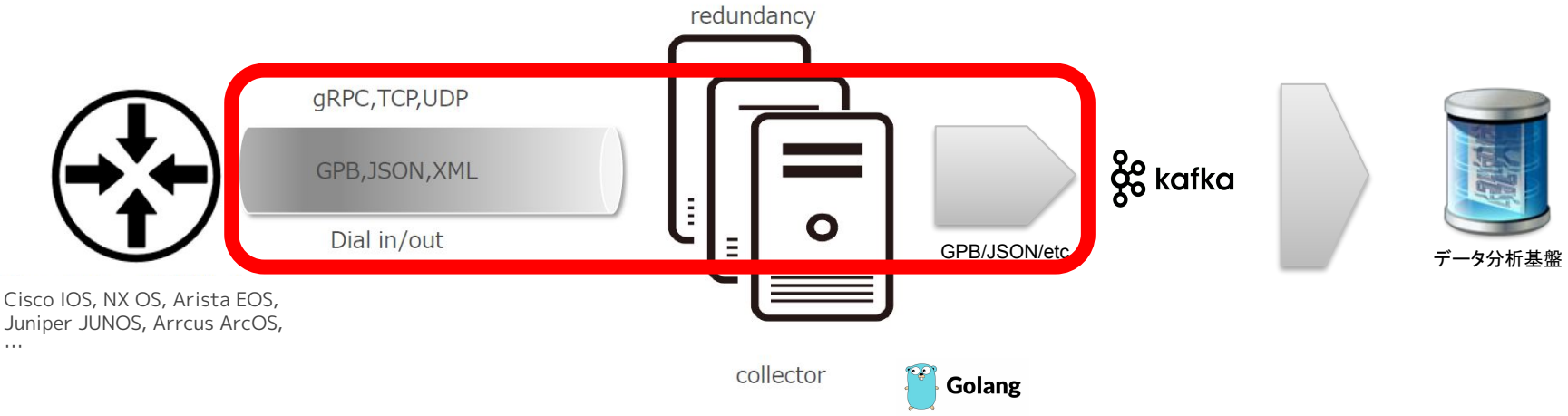


最低限必要だったこと④

SoftBank

トラフィックをさばく仕組み

- Telemetry取得・転送用のNW（広帯域）を用意し大規模データ取得を可能に

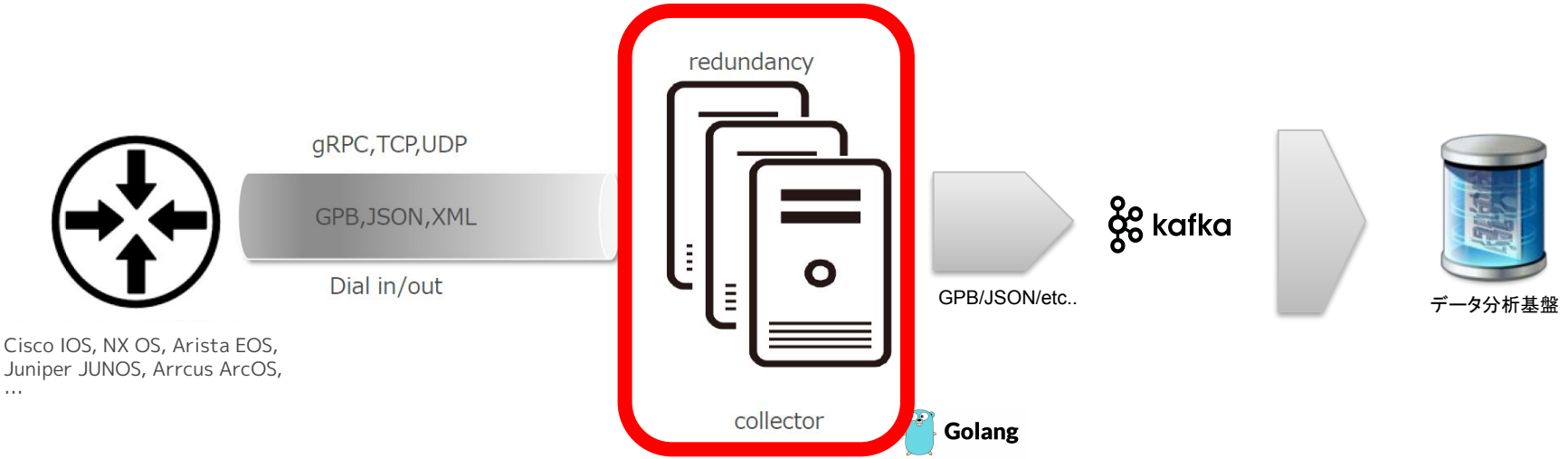


最低限必要だったこと⑤

SoftBank

各コンポーネントの冗長可

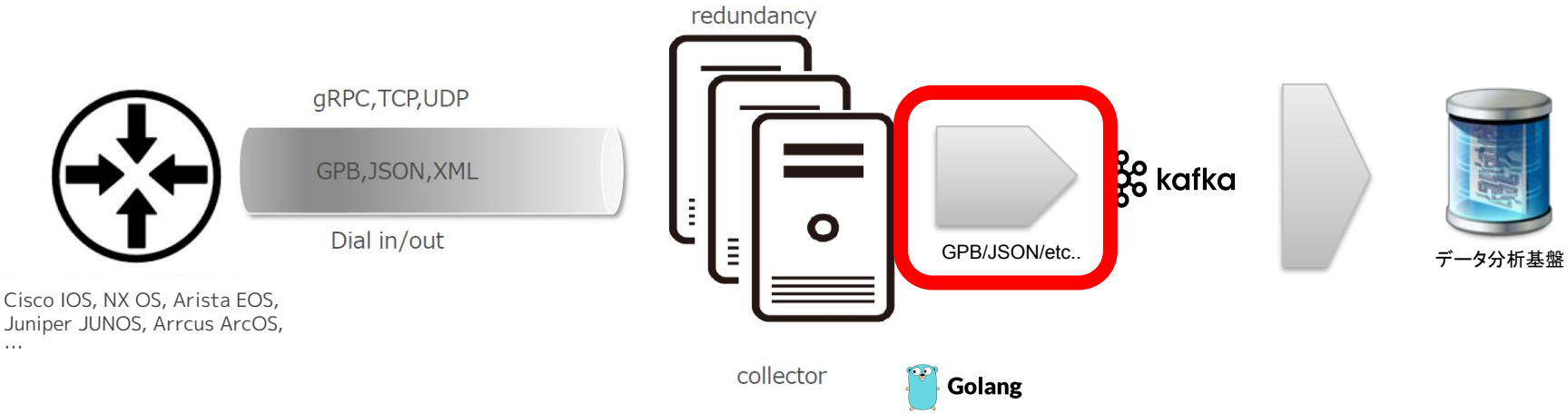
- 取得先が商用ノードなので一定の対障害性が必要



SoftBank

後段へのデータ転送

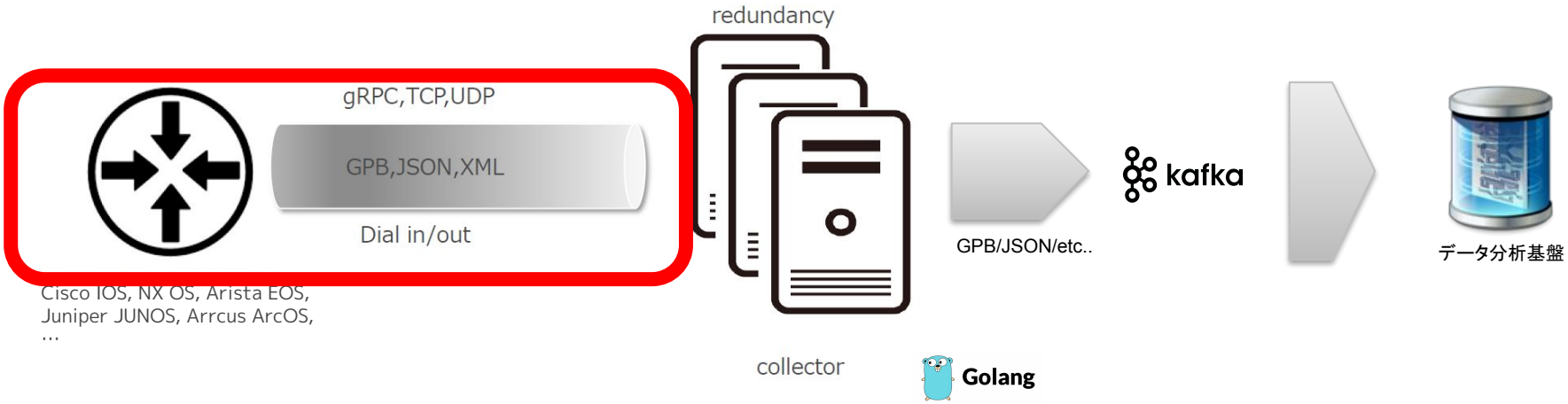
- データを処理し、後段の分析基盤へ漏れなく転送
- kafkaの転送管理とか再送処理とか



SoftBank

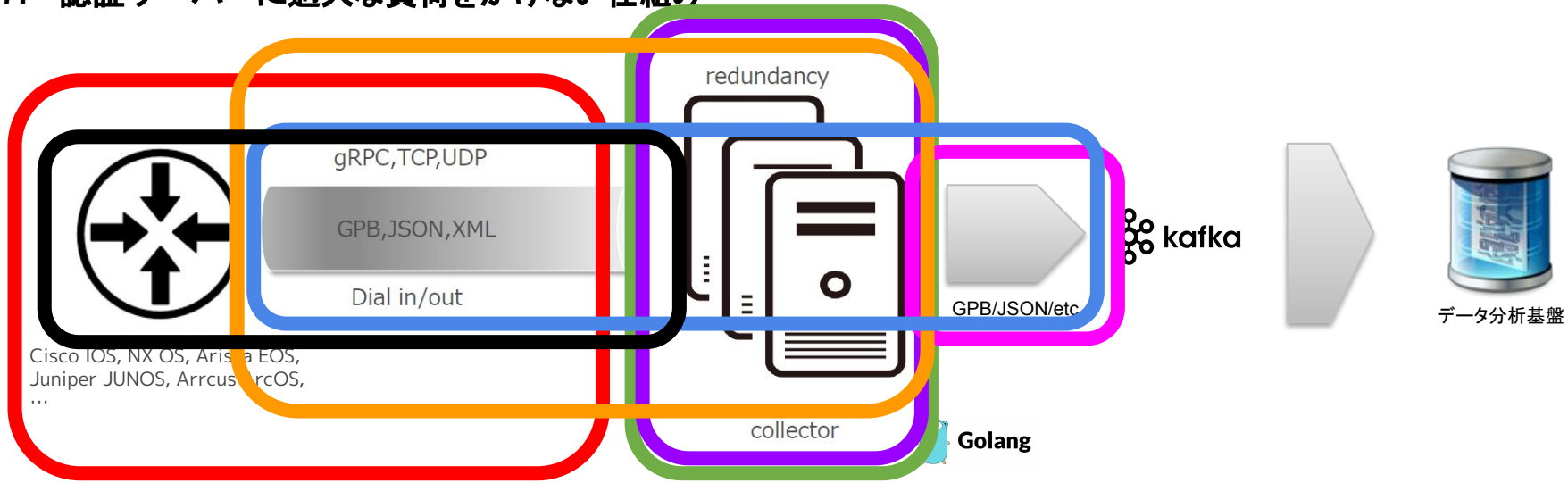
認証サーバーに過大な負荷をかけない仕組み

- 認証情報/セッションのキャッシュ化など



まとめ

1. 複数ルータ対応
2. ルータ側 down / S.O. 後のデータ取得自動再開
3. 様々な用途のデータを一様に管理
4. トラフィックをさばく仕組み
5. 各コンポーネントの冗長可
6. 後段へのデータ転送
7. 認証サーバーに過大な負荷をかけない仕組み



コレクタ開発のメリット

SoftBank

	独自コレクタ	OSSコレクタ
導入容易性	△ 小規模開発ではあるが スクラッチ開発が必要	○ 大手通信機器ベンダーから OSS公開あり
運用保守性	◎ 開発チームによるサポートあり	✕ 原則サポートなし
システム連携 (後続)	◎ 後段システムの構成にあわせた 柔軟な対応が可能	△ OSS内で定められた手法のみ可能
ベンダー制約	◎ 仕様が分かれば 全ベンダ対応可能	✕ コレクタの仕様に依存

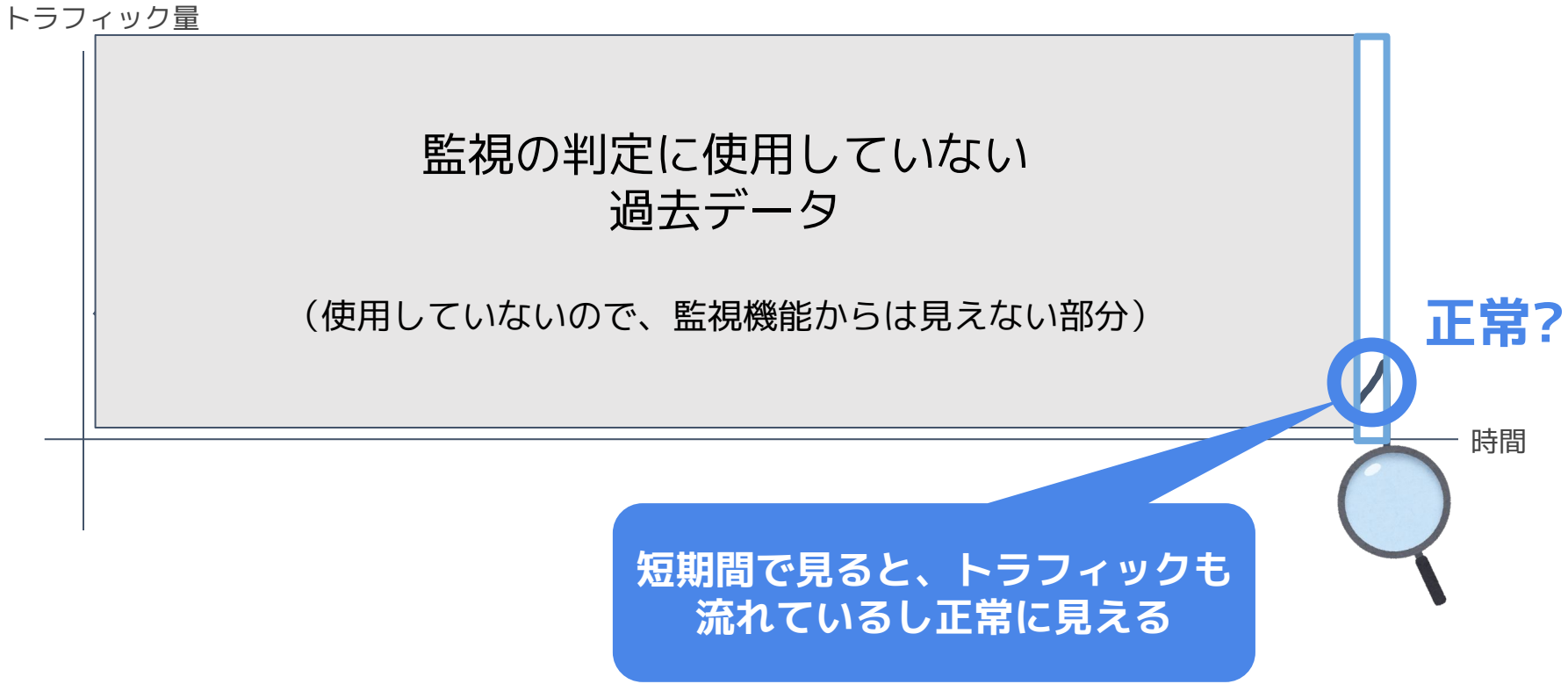
生みの苦しみを越えてしまえば自由な運用が可能

傾向異常検知

短期間の視点だけではなく

SoftBank

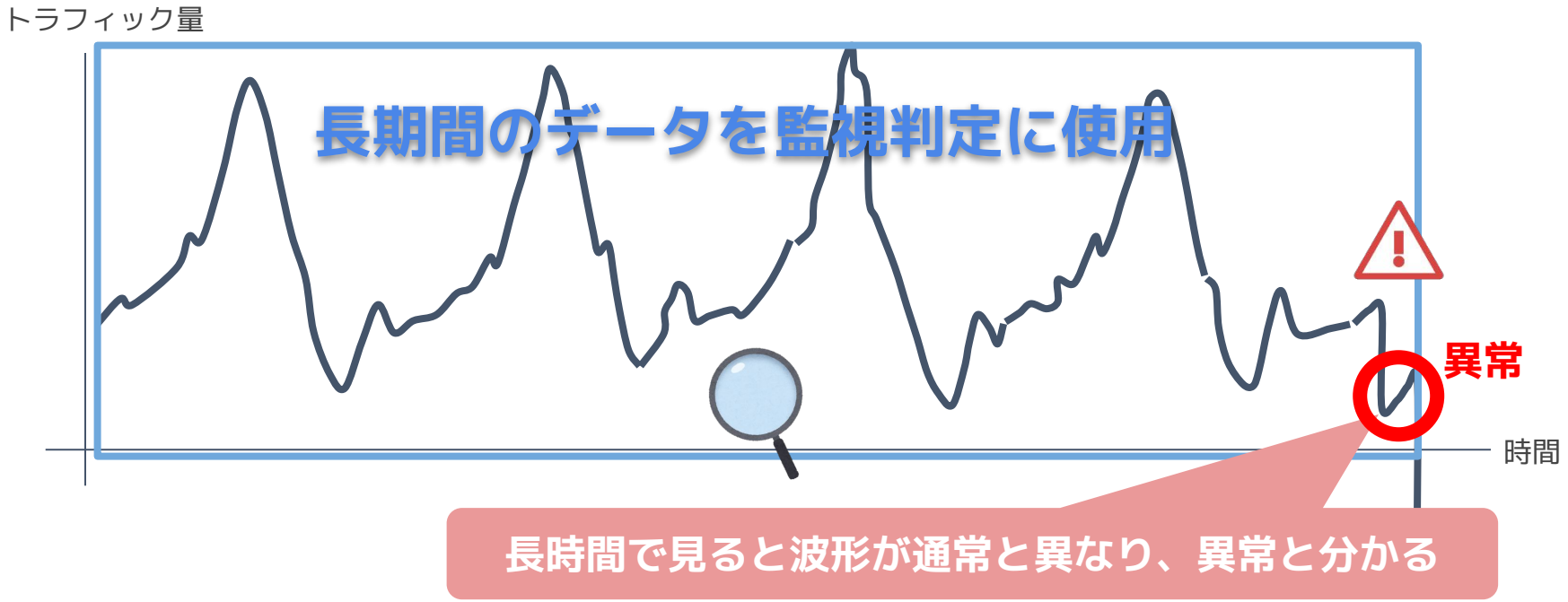
ある日のトラフィック流量を示すグラフにて



長期間の視点も必要

SoftBank

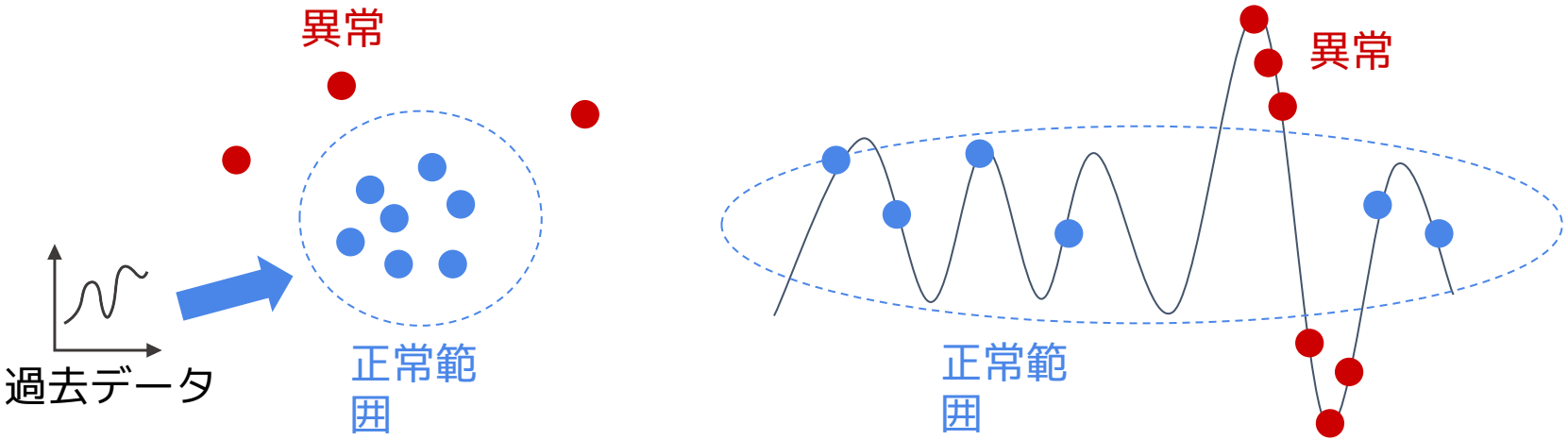
ある日のトラフィック流量を示すグラフにて



傾向異常検知が必要

傾向異常検知における異常の定義

傾向異常検知における異常のパターンは多様で、未知のパターンも含む
　→決まったパターンや固定の閾値での監視が困難
　→よって、過去データの分析結果から正常な状態を設定し、そこから外れた状態を異常として判定する
　　（外れ値検知）

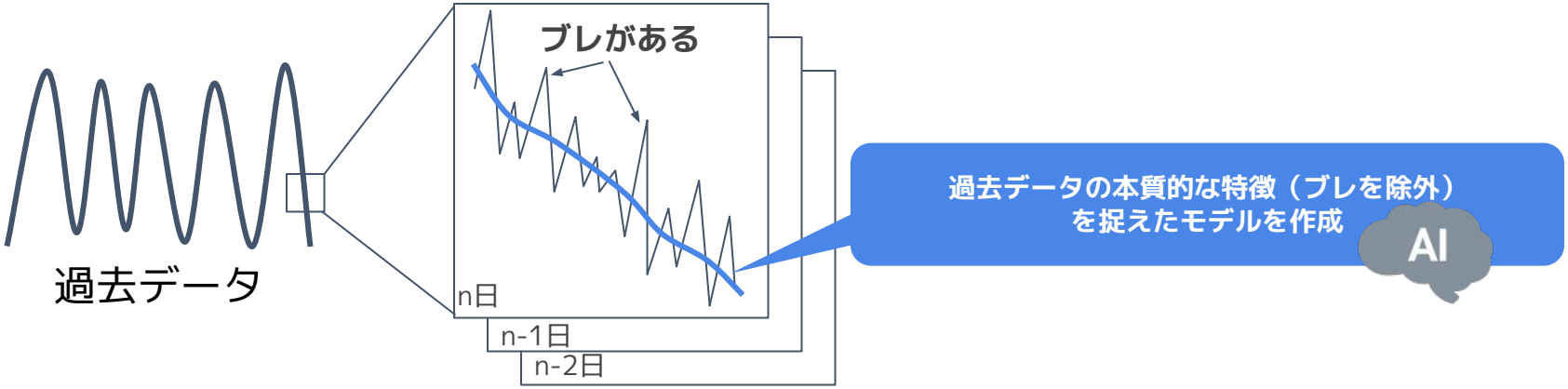


外れ値検知のイメージ

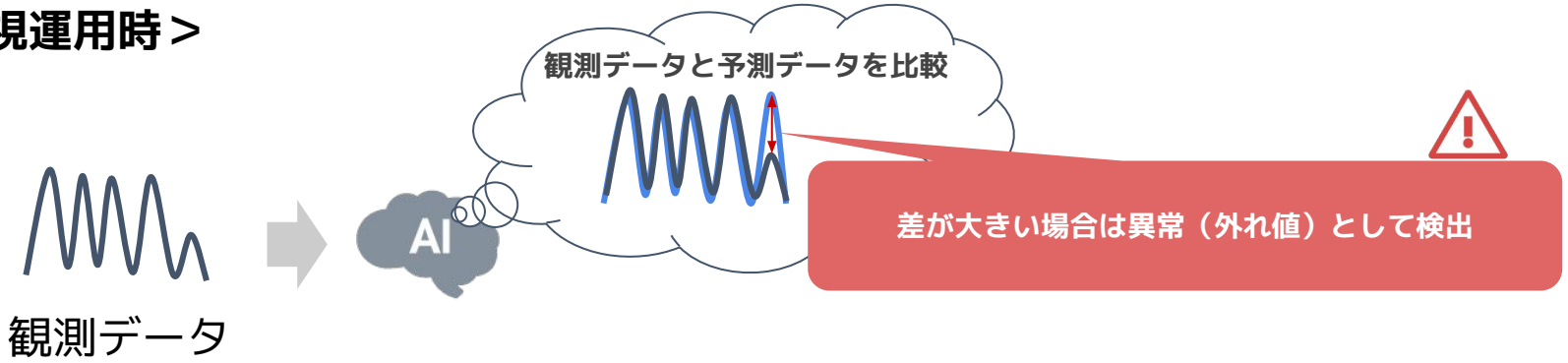
傾向異常検知の導入（周期特性）

SoftBank

＜学習時＞



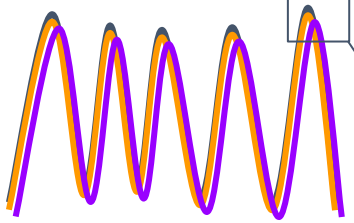
＜監視運用時＞



傾向異常検知の導入（相関特性）

SoftBank

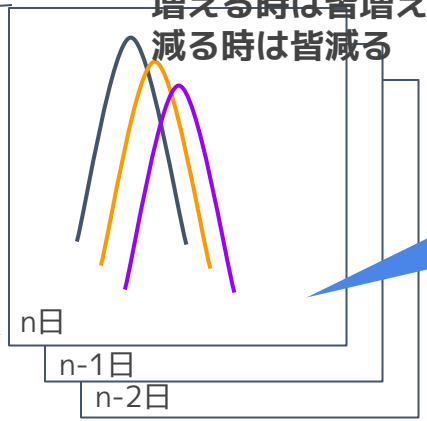
＜学習時＞



過去データ

増える時は皆増える
減る時は皆減る

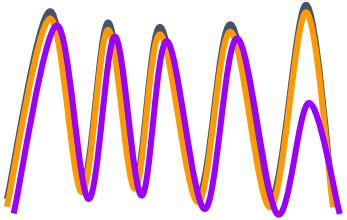
=相関特性がある



過去データの相関関係について
本質的な特徴（ブレを除外）を捉えたモデルを作成

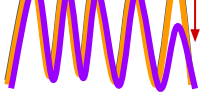
AI

＜監視運用時＞



観測データ

観測データと予測データを比較



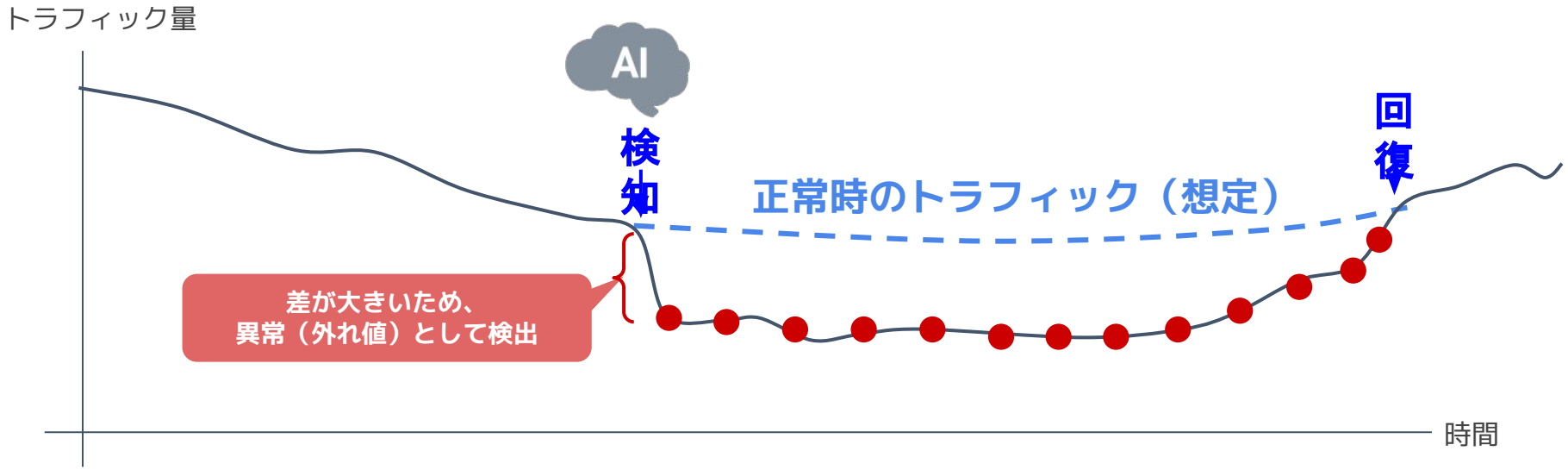
AI

相関関係が崩れている場合は
異常（外れ値）として検出

傾向異常検知が活きるパターン例

NW運用の中で、稀に発生する障害

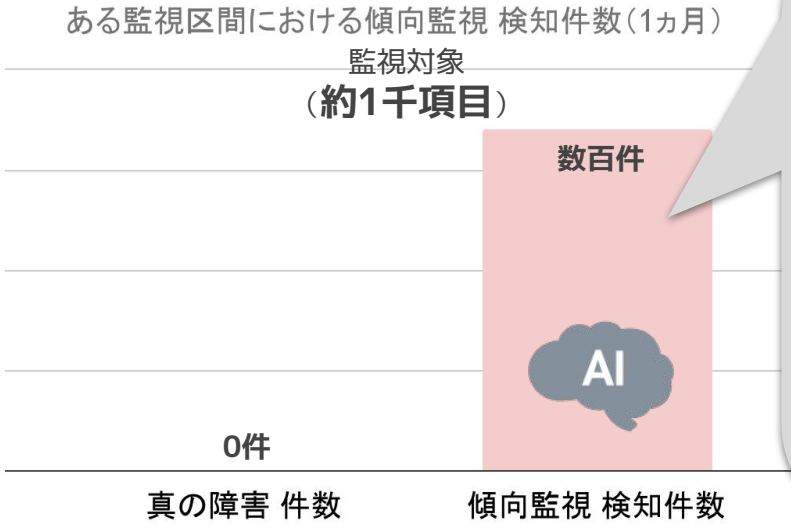
- **スループットが低減**（ex.5割減）
- しかし、完全に通信断になる訳ではない（半死状態）ので、ping監視などでは気付きづらい



しかし、この仕組みだけでは課題が

SoftBank

＜課題＞
真の障害の発生件数（数年に1件か、それ以下）と比較し、
傾向異常検知の検知件数が高い



**少しでも怪しかったら
通知するポリシーだとして
も。。。**

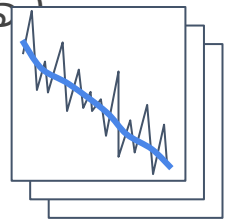
**通知先がオペレータ（人間）なの
で
検知数が多い**

**確認
しきれないよ。。**

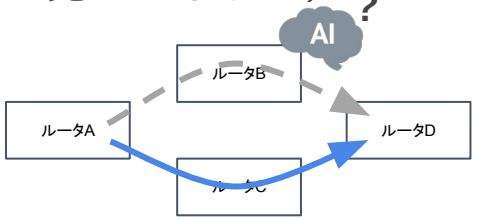
なぜ検知件数が多いのか？

- 周期特性や相関特性は、常に一定ではない（誤差が多分に含まれる）

本質的特徴を学習するも
完璧ではない



- NW運用により、通信経路の変更が散発的に発生
（NWサービスは正常だが、事前学習した正常範囲から外れたように見えてしまう）



- トレンド変化への追従のために定期的な再学習を行うが、良い学習データが得られるとは限らない（※上記の通信経路変更などの影響etc）

良くない
学習データ

他にも色々、要因がある。。

それでも傾向異常検知には、まだまだ課題あり

SoftBank

監視対象の全体母数が多い（数千項目以上）ため、低確率でも誤検知が**月に10〜数十件発生**してしまう場合がある

監視メトリックのサマリ化
（ex.複数監視項目の合算値）

→周期性は安定するが、特定区間の異常が薄まってしまい鈍感になってしまう

異なる種別のメトリックを複合的に判断

→関連性がある項目を組合せないと意味がなく、未知の異常ではどの組合せに意味があるのか不明

LLMを使って、誤検知と思われるものを分類したいが、現在試行中

＜議論したい点＞

傾向異常検知を導入している組織は他にもあると思う。
機能や運用での工夫点など、是非議論したい

まとめと議論のポイント

SoftBank

＜リアルタイム検知＞

センサパスに関する課題

- 一部のOSでは**センサパスを機器側に設定**
- **適切なセンサパスの探し方**

データ量の肥大化

- 本来は**多数のパスからデータ蓄積**をしたい。
  → 障害時にクリティカルなパスを確認

NW機器自体のVerUP追従

- バージョンによる**取得パス**の違い

データ取得間隔に関する課題

- **数秒のスパンで見えるもの**/実例があるか。

＜傾向異常検知＞

傾向異常検知を導入している組織は他にもあると思う。
機能や運用での工夫点など、是非議論したい

Appendix

End of Slide