# Chapter 11
# A Perspective on the Standardization of Autonomic Detection of Service Level Agreement Violations

**Jéferson Campos Nobre**
*University of Vale do Rio dos Sinos, Brazil*

**Lisandro Zambenedetti Granville**
*Federal University of Rio Grande do Sul, Brazil*

## ABSTRACT

*Service level agreements (SLAs) allow networked services established between providers and their customers to operate according to the conditions defined in the SLA. Measurement mechanisms can be used to support SLA monitoring. However, these mechanisms are expensive in terms of resource consumption. In addition, if the number of SLA violations at any given time is greater than the available measurement sessions, some violations will likely be missed. The current best practice is to observe just a subset of network destinations based upon the expertise of a few human administrators. Such observation mode is error prone, reactive, and scales poorly. Such practice can lead to SLA violations being missed, which hampers the reliability of the SLA monitoring process. In this context, the use of autonomic network features can improve such processes, especially when these features are deployed in a decentralized manner. The use of these autonomic features is described in RFC 8316. The authors expect that such a document can lead to better SLA monitoring tools and methods.*

## INTRODUCTION

Communication requirements of distributed services running on top of an IP-based infrastructure have become increasingly demanding. Some examples are HealthCare applications (eHealth) and data-intensive science applications (eScience). The provisioning of such services with the adequate level of quality, as typically documented in the Service Level Specification (SLS) that pertains to the Service Level Agree-

DOI: 10.4018/978-1-5225-7146-9.ch011

ment (SLA), is conditioned by the accommodation of requirements that are usually expressed in terms of metrics, such as inter-packet delay variation, packet loss or latency. Such requirements usually lead to the definition of Service Level Objectives (SLOs) that must be met. Those SLOs are part of SLAs that define a contract between the provider and the consumer of the service.

Performance requirements can be employed effectively by both service providers and customers. In this context, SLOs reflect a service-level guarantee that the consumer of the service can expect to receive. Likewise, the provider of a service needs to ensure that the service-level guarantee and associated SLOs are met. When such SLOs are not met, SLAs usually include financial or other penalties, possibly with the risk of cancelling the deal. Besides, an adequate support of SLAs also improves the commercial reputation of the service provider, considering prospective customers.

The detection of SLA violations is based on the idea of identifying deviations from the contracted SLOs. In order to identify these deviations by using active measurements, it is necessary to have measurement sessions activated on key end-to-end network destinations. However, such activation is expensive in terms of resource consumption (both human and computational), let alone the amount of monitoring traffic that may jeopardize the performance of network devices and network bandwidth. Since a better monitoring coverage requires more active sessions, it increases the amount of consumed resources. On the other hand, enabling the observation of just a subset of all network flows decreases the resource consumption, but it can lead to insufficient coverage.

The decision about how to place measurement sessions is an important management issue, since it impacts the SLA monitoring coverage. The goal is to obtain the maximum coverage with a limited amount of measurement overhead. Specifically, the goal is to maximize the number of SLA violations that are detected with a limited number of resources. In this context, a feasible approach would be to add up the service levels observed across different path segments. This allows the decomposition of a large set of end-to-end measurements into a much smaller set of segment measurements. However, some end-to-end service levels cannot be determined by an additive approach. Some examples of metrics that are inadequate for additive approaches are end-to-end jitter and mean opinion scores, thus they must be measured end-to-end (Nobre, Granville, Clemm, & Prieto, 2018a).

Often, the current best practice for activating measurement sessions within a provider's network consists in relying on the network administrator's expertise to determine which destinations to select to activate the corresponding monitoring sessions. This practice has major shortcomings. Indeed, such practice assumes high dynamics and increases the complexity of network environments and delivered services. In order to provide solutions that better suit such dynamics and complexity, network-wide management solutions can be employed. A network-wide control of network devices can improve their abilities to accomplish management tasks. For example, a distributed network management algorithm can be used to allow that some devices provide additional resources for the execution of management tasks by other devices. This can be useful when either the computational load is not equally distributed among the network devices or there is heterogeneity in the computational resources of network devices. In this context, the global capability of the devices in a network can be greater than the sum of the capabilities of each device.

There is no standard solution for a distributed and autonomic detection of SLA violations. Current solutions are restricted to ad hoc scripts running on each node to automate some administrator actions. Network management researchers have investigated how to overcome analogous shortcomings considering different scenarios and management tasks. There are some proposals for passive probe activation - e.g., DECON (Di Pietro, Huici, Costantini, & Niccolini, 2010) and cSamp (Sekar, Reiter, Willinger,

Zhang, Kompella, & Andersen, 2008) - but these do not focus on autonomic features. In this context, RFC8316 (Nobre, Granville, Clemm, & Prieto, 2018a) describes a use case for the autonomic detection of SLA violations considering the use of Peer-to-Peer (P2P) techniques. The present chapter presents an overview of RFC8316, including the network management context which yielded its publication.

This chapter is organized as follows: In the Background Section, the authors present the fundamental concepts considering network measurements and P2P techniques in network management. In RFC8316 Section, the authors detail the context and the perspective of this document. In the Future Research Directions Section, the authors present what future holds for autonomic detection of SLA violations. Finally, concluding remarks are provided in the Conclusion.

## BACKGROUND

### Network Measurements

Several mechanisms can be used to enable network measurements, e.g., in-band/in-situ (Operations, administration and management) OAM (Norton, 2016). In general, these mechanisms are classified according to the injection of measurement traffic by the mechanisms themselves. This usually leads to two types of network measurement mechanisms: passive and active measurement mechanisms. Passive measurement is realized, for example, inside network devices through the use of packet sniffers. On the other hand, active measurement is deployed by means of active probes hosted along the network which inject synthetic (i.e., artificially generated) traffic and compute the current network performance. In addition, there are mechanisms that cannot be classified according to these two categories, and they are usually called hybrid mechanisms. In this subsection, the authors first cover some background on passive, active, and hybrid mechanisms.

### Passive Measurement Mechanisms

In passive measurement, network conditions are said to be checked in a non-intrusive way, because no monitoring traffic is created by the measurement process itself. Passive measurement data can be used for a variety of purposes. Passive measurement is realized, for example, inside network devices when they observe the traffic flows that cross the device. In the context of IP Flow Information Export (IPFIX), several documents were produced to define how to export data associated with flow records, i.e., data that is collected as part of passive measurement mechanisms, generally applied against flows of production traffic (e.g., (Claise, Trammell, & Aitken, 2013). In addition, it is possible to collect real data traffic (not just summarized flow records) with time-stamped packets, possibly sampled (e.g., (Duffield, Chiou, Claise, Greenberg, Grossglauser, & Rexford, 2009), as a means of measuring and also inferring service levels.

Flows can be defined as unidirectional sequences of packets that pass through a network device which are grouped according to some common properties. These properties can consider several packets fields, such as source/destination IP address and source/destination port number, layer 3 protocol identifier, Type of Service (ToS), and size (aggregated number of bytes). In addition, other information, such as source/destination Autonomous System (AS), and input/output interfaces can also be used to define flows. Representation of flow data must be uniform/homogeneous, as well as communication means

to exchange such data between the network and the collection points (Claise, 2008). There are several protocols used to enable flow data production and exchange.

The IETF IP Flow Information eXport (IPFIX) Working Group has released several documents describing a protocol, based on the version 9 of NetFlow (Claise, 2008). Some enhancements in different domains (e.g., congestion-aware transport protocol and built-in security) were incorporated in the IPFIX protocol. Furthermore, IPFIX adopts an improved use of record templates through more precisely defined record items and measurable values. Unlike NetFlow, IPFIX requires Stream Control Transport Protocol (SCTP) (Dreibholz, Rathgeb, Rungeler, Seggelmann, Tuxen, & Stewart, 2011) to transport data. The use of SCTP provides a reliable transport and prevents congestion. Figure 1 shows the IPFIX logical model (which is based on the NetFlow logical model) as an example of a passive measurement model. In such model, metering exporters hosted in network elements (e.g., routers and switches) gather flow data and export IPFIX records to configured receivers, i.e., collectors (or collecting points).

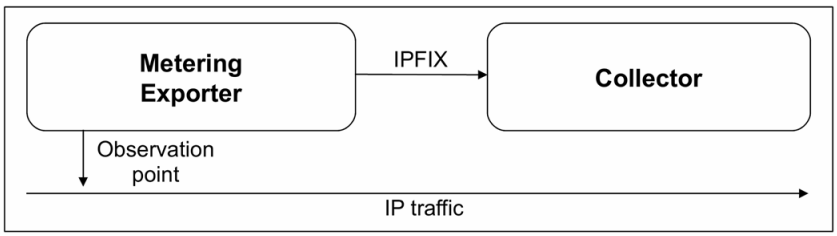
## Active Measurement Mechanisms

Active measurement mechanisms can be used to monitor SLOs and the health of a network as a whole. Such mechanisms inject synthetic traffic into specific network paths to measure the network performance in terms of, for example, delay, loss, jitter, and packet/frame loss. A well-defined injection of such traffic is usually called a measurement session. Active measurement mechanisms can be employed in different contexts, such as pre-deployment service validation and live network-wide SLA monitoring.

The generation of synthetic traffic and its computation to provide measurements results accordingly is usually performed according to an architecture that is comprised of two hosts with specific roles, a sender and a responder, also collectively known as (active) measurement probes. The exchange of packets between probes is usually defined by two inter-related protocols: a control protocol, used to initiate and control measurement sessions and to fetch their results, and a test protocol, used to send single measurement packets along the network path under test. Measurement support at the responder end may be limited to a simple echo function. There are several protocols used to enable active measurement.

Cisco Systems defines the SLA protocol (also known as IPSLA) which is described in an IETF informational RFC (Chiba et al., 2013). This widely deployed protocol measures service levels related to data link and network layers and it also emulates characteristics of different applications, both considering one-way and two-way metrics. The IPSLA logical model consists essentially of a sender and a responder, i.e., measurement probes. The protocol consists of two distinct phases: the control phase and the measurement phase. The control phase forms the base protocol, which establishes the identity of

*Figure 1. Passive Measurement Model (IP Flow Information eXport)*
*Source: Claise, 2008*



the sender and provides information for the measurement phase. The measurement phase is comprised of a sequence of measurement-request and measurement-response messages (test messages). Figure 2 shows the logical model used by IPSLA.

The IETF IP Performance Metrics (IPPM) Working Group has proposed open active measurement mechanisms that allow the exchange of packets to produce one-way and two-way metrics. These mechanisms are called One-Way Active Measurement Protocol (OWAMP, RFC4656) (Shalunov, 2006) and Two-Way Active Measurement Protocol (TWAMP, RFC5357) (Hedayat, 2008), respectively. The O/TWAMP mechanisms consist of two inter-related protocols: a control protocol, used to initiate and control measurement sessions and fetch their results, and a test protocol, used to send single measurement packets along the Internet path under test. Control protocol is performed by the control-client (requests, starts, and ends test sessions) and server (manages test sessions); and the test control is then executed by the sender (sending endpoint) and session-receiver/reflector (receiving endpoint).

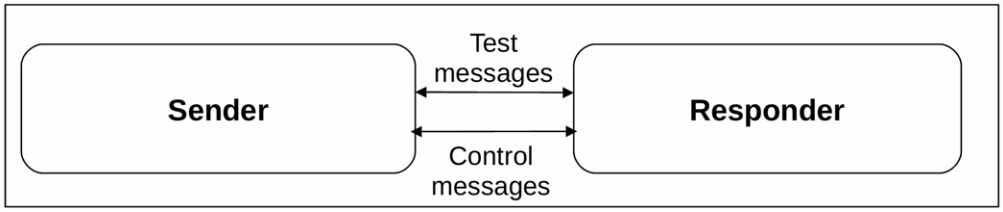
## Hybrid Measurement Mechanisms

The notions of active and passive measurement mechanisms are well-established. On the contrary, hybrid measurement mechanisms, which combine active and passive techniques, are fuzzier, and their definition varies depending on the standardization body. These hybrid mechanisms aim at providing some of the benefits of both active and passive mechanisms without embarking their disadvantages. In this context, the metrics produced by hybrid measurement usually have distinct properties, and they consume different amounts of resources for delivering the results. For example, in terms of delay, hybrid methods could provide results with better accuracy than passive methods, but without the network cost of active methods.

The IETF IPPM WG proposed a categorization for hybrid measurement mechanisms (Morton, 2016). The first type, Hybrid Type I, is based either on the augmentation or modification of the stream of interest, or the use of methods that modify the treatment of such stream. The second type denoted as Hybrid Type II, defines the use of mechanisms that use two or more different streams of interest with some degree of mutual coordination to collect both active and passive metrics and enable additional joint analysis. Some use of hybrid measurement mechanisms is also defined as spatial metrics and methods (Morton, 2016).

The IPv6 Option Header for Performance and Diagnostic Measurements (PDM) Internet-Draft (Elkins & Hamilton, 2017) proposes the addition of fields dedicated to the measurement of user streams. The measured stream has unknown characteristics until it is processed to add the PDM Option header. The use of PDM intends to have a minor effect on the measured stream and other streams in the network when it is added to network interfaces. Considering the IPPM classification, this is a Hybrid Type I method, having at least one characteristic of both active and passive mechanisms for a single stream of interest. The

*Figure 2. Active Measurement Model (Cisco Service Level Assurance protocol)*
*Source: Chiba, Clemm, Medley, Salowey, Thombare, & Yedavalli, 2013*



Alternate-Marking Method for Passive and Hybrid Performance Monitoring (Chen, Castaldelli, Mirsky, Mizrahi, Fioccola, Capello, & Cociglio, 2018) is a Hybrid Type I method as well. In this method, the stream is measured and time-stamped during that process to deliver network metrics. Thus, data packets are marked by different blocks of markers that change one or more bits of packets without altering either the normal processing in the network or adding delimiting packets regarding the measurement traffic.

This subsection only discusses the passive, active and hybrid measurements mechanisms. Readers may want to look at the survey performed by Nobre et al. (2018a) for more details.

## P2P-Based Network Management

There is substantial research on models that addresses the structure of interactions required to execute network management tasks (dos Santos, Famaey, Schönwälder, Granville, Pras, & De Turck, 2016). In these models, various forms of decentralization (i.e., distribution) are used to produce, access, and store management data. In the traditional centralized model, a single management station typically controls the whole managed infrastructure. Scalability issues of the centralized model motivated intense research on Distributed Network Management (DNM) alternatives. Some management literature classifies the various flavors of DNM solutions (dos Santos, Famaey, Schönwälder, Granville, Pras, & De Turck, 2016). A possible approach to decentralize the execution of management tasks is to employ P2P techniques. Such techniques have proven to be adequate for different kinds of applications. Therefore, P2P techniques may also be successfully used for DNM. Figure 3 presents a general view of the P2P-Based Network Management (P2PBNM) model.

*Figure 3. P2P-Based Network Management Model*
*Source: Nobre, Granville, Clemm, & Prieto, 2012*

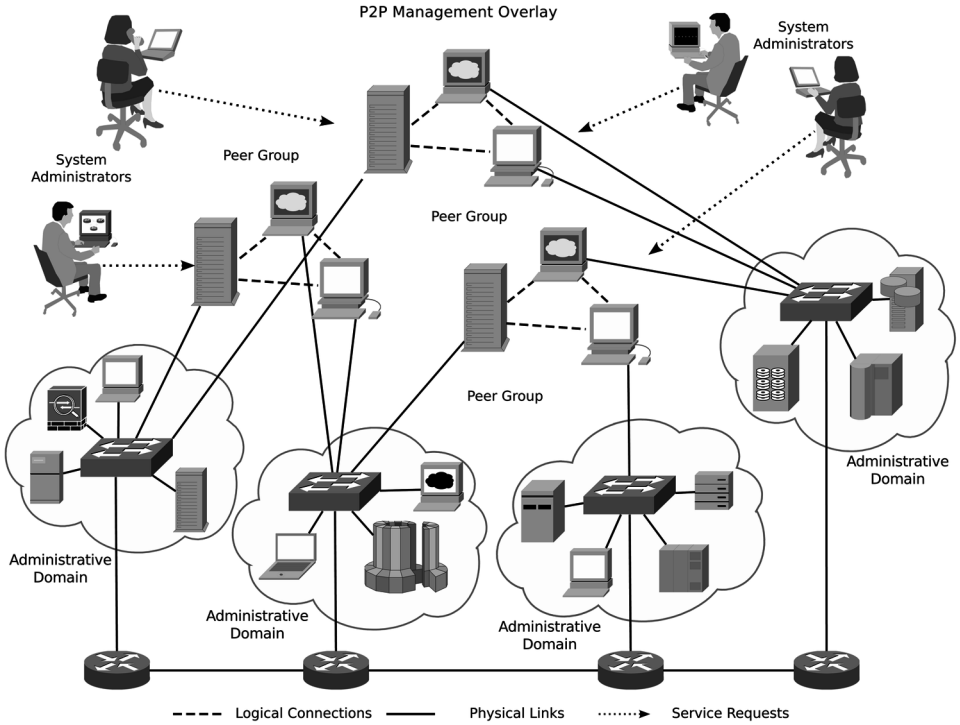


Figure 3 presents a P2P overlay in which shared computational resources are used to perform management tasks. The choice of protocols used to build a P2P management overlay differs significantly among P2PBNM research initiatives and prototypes. Some initiatives reuse well-established P2P protocols in order to reuse the properties of these protocols, which were already described in the literature. For example, the *Cyclon* protocol (Voulgaris, Gavidia, & Van Steen, 2005) is used for deploying management overlays. In addition, the reuse of P2P protocols eases the development of a P2PBNM system since the focus can remain on the management tasks (and not in building new P2P protocols). On the other hand, some initiatives built a P2P protocol from the scratch focusing only on the required features/properties to make the management overlay operational. A P2P protocol designed specifically for DNM does not add "compulsory" overheads needed to address requirements of general purpose P2P systems. Therefore, the P2PBNM system efficiency can be increased.

The approach to distribute management tasks also varies, depending on the nature of the P2PBNM approach. One possibility is to use Service-Oriented Architecture (SOA, Jones, 2005). When using SOA, management peers perform management tasks through management services. In this context, the result of these services is the execution of a management task. In general, these services are requested by system administrators (as shown in Figure 3) or automation procedures (which can be hosted either inside the peers themselves or even in a centralized entity). The software portion that is responsible for delivering management services is usually known as a management component. The spectrum of the technology used by these components is very broad, e.g., from simple monitoring probes to complex autonomic policies interpreters.

P2P techniques may also be a valuable tool to enable inter-domain management (Fiorese, Simões, & Boavida, 2009). P2PBNM systems typically use Application Layer Routing (ALR, RFC5693) (Burger & Seedorf, 2009) as their main message passing resource and ALR adapts more easily to administrative domain boundaries. In this context, logical connections among the peers are mapped into physical links. Figure 3 illustrates a scenario where participating peers (of peer groups) spread over different administrative domains; logical connections among peers are represented by dashed lines. Management entities in traditional management rely on the IP routing to communicate with one another. Thus, the definition of new routes is rigidly produced by current routing protocols. Furthermore, boundary boxes (e.g., application proxies, packet filters) break the end-to-end IP forwarding paradigm. The use of ALR can overcome network layer connectivity issues or, at least, optimize connectivity using information from the network layer (Burger & Seedorf, 2009).

## RFC8316

RFC8316 (Nobre, Granville, Clemm, & Prieto, 2018a) depicts the use case of a service provider that needs to monitor the network it operates to detect SLA violations by using active measurements mechanisms, while limiting human intervention. The goal is to conduct the measurements in an effective manner to maximize the percentage of detected SLA violations with resource restrictions. This section describes RFC8316 and its context.

## Background

There is a minimum set of properties which define an Autonomic System: automatic, i.e., it can "self-control its internal functions and operations"; adaptive, i.e., it can change its "configuration, state and

functions"; and aware, i.e., it can "monitor its operational context". In addition, various definitions of an Autonomic System have been given, such as self-CHOP and MAPE-K. The application of autonomic systems to the complete network lifecycle (e.g., installation, commissioning, operating) is usually called Autonomic Networking (AN).

There has been substantial research on autonomic features related to network management solutions. The application of Autonomic Computing (AC) principles in network management, referred to as Autonomic Network Management (ANM), has been proposed as a means to address some issues faced by traditional network management, such as controlling highly dynamic environments as in ad hoc networks.

There were several antecedents to AN, for example, the use of artificial intelligence in network management, Self-Organising Networks, Declarative Policies, etc. Using the term "AN", Strassner (2004) presented a tutorial during a Network Operations and Management Symposium (NOMS) conference. This was one of the first traceable academic mentions of AN. Since 2004, some initiatives have been publicized, and they discussing different attributes of AN (e.g., Mortier & Kiciman, 2006; Agoulmine, Balasubramaniam, Botvich, Strassner, Lehtihet, & Donnelly, 2006).

AN was the focus of several research projects over the last decade. Some examples of such projects are AN Architecture (ANA) (Bouabene, Jelger, Tschudin, Schmid, Keller, & May, 2010), Generic ANA (GANA) (Chaparadza, 2009), etc. In this context, AN is usually addressed by several publications which have been presented over the years in the Network Management conferences (e.g., IM, NOMS, and CNSM). In the meantime, other networking technologies gained momentum, such as Software-Defined Networking (SDN) and Network Functions Virtualization (NFV). Finally, some new notions have been developed, and they somewhat overlap with AN, such as cognitive networking and intelligence-driven networking. However, there is a lack of successful deployment cases and standardization regarding AN.

The Network Management Research Group (NMRG) from the Internet Research Task Force (IRTF) also started to discuss AN. The first AN-related discussions were held during workshops, called "Autonomics for Network Management @ NMRG". There were also related efforts conducted by other IETF WGs and IRTF RGs, such as Simplified Use of Policy Abstractions (SUPA), Home Networking (HOMENET), Software Defined Networking Research Group (SDNRG), Network Functions Virtualization Research Group (NFVRG), Interface to the Routing System (I2RS), among others. In addition, some new groups have been proposed, such as Network Machine Learning Research Group (NMLRG) and Intelligent Defined Networking (IDN).

Some important outcomes of NMRG meetings (and its work in general) are related to autonomic networking. First, 2 RFCs were produced in this context, RFC7575 and RFC7576. The first document defines a common language and outlines design goals for autonomic functions (Behringer, Bjarnason, Jiang, Carpenter, Pritikin, Ciavaglia, & Clemm, 2015). The second document details a problem statement and a general gap analysis for an IP-based Autonomic Network that mainly relies upon distributed network devices (Jiang, Carpenter, & Behringer, 2015). Finally, there was the Use Cases for Autonomic Networking (UCAN) Bird of Feathers (BoF) session held during the IETF 90 meeting, and which attracted a large audience.

The UCAN BoF led to the creation of the ANIMA (Autonomic Networking Integrated Model and Approach) WG. With a focus on professionally-managed networks, the main goal of ANIMA is to produce the specification of a minimum set of reusable infrastructure components to support autonomic interactions and use cases. In this context, the definition of autonomic networking by ANIMA is "a system of autonomic functions that carry out the intentions of the network operator without the need for detailed low-level management of individual devices" (refer to Chapter 4 "*ANIMA: Secure Autonomic Network*

*Infrastructure*"). However, some of the Internet-Drafts presented during the UCAN BoF session have not been hosted by ANIMA as WG items, and the corresponding discussion was therefore progressed within NMRG. The Autonomic Networking Use Case for Distributed Detection of Service Level Agreement (SLA) Violations Internet-Draft was one of such "left aside" Internet-Drafts.

## Use Case Description

The use case and the solution described in RFC8316 address an important practical issue, the detection of SLA violations. They are intended to provide a basis for further experimentation that can lead to solutions for wider deployment. Solutions that allow the service provider to monitor and troubleshoot the underlying communication infrastructure are crucial. In the RFC8316, the authors considered that this provider has a restricted resource budget with regard to SLA monitoring coverage, specifically in terms of the number of measurements that can be conducted concurrently and possibly the total amount of measurement traffic that is forwarded across the network.

The activation of measurement sessions is done through several steps. First, it is necessary to collect traffic information in order to grasp the traffic matrix. Then, the administrator uses this information to infer the best destinations to activate measurement sessions (i.e., the ones most prone to have SLA violations). After that, the administrator activates sessions on the chosen subset of destinations, taking the available resources into account. This practice has major shortcomings. It does not scale well and usually covers only a fraction of the network flows that should be observed. Network management researchers have investigated how to overcome similar shortcomings considering different scenarios and management tasks (Samaan & Karmouch, 2009). In order to provide solutions that better suit these scenarios, network-wide management solutions can be employed.

A network-wide control of network devices can improve their ability to accomplish management tasks. For example, a distributed network management algorithm can be used to allow some devices to provide additional resources for the execution of management tasks by other devices. This can be useful when either the computational load is not equally distributed among the network devices or there is heterogeneity in the computational resources of network devices. In this context, the global capability of the devices in a network can be greater than the sum of the capabilities of each device. In addition, the solution needs to be dynamic, being able to cope with network conditions that may change over time, and embeddable, considering network devices that control the deployment of measurement mechanisms.

The goal of an autonomic solution in the present use case is to conduct the measurements in a smart manner that ensures that the network destinations are broadly covered and that the probability of detecting SLA violations is improved. Clearly, static (i.e., defined a priori) solutions will have severe limitations. While at any given time, the number of measurements in progress is limited, it is possible for a device to change the destinations to measure over time. Thus, an autonomic solution is needed so that network measurements are automatically orchestrated and dynamically reconfigured from within the network. This can be exploited to achieve a balance of eventually covering all possible destinations using a reasonable amount of "sampling" (i.e., coverage versus resource consumption) where measurement coverage of a destination cannot be continuous.

A solution for the detection of SLA violation should focus on the use of measurement resources on destinations that are more likely to incur a violation (Nobre, Granville, Clemm, & Prieto, 2012). In any case, the solution still needs to spend (possibly) fewer resources on destinations that are more likely to be respecting SLOs. However, when a probe first comes online, it has no information about which

measurements are more critical than others. At the same time, human administrators should not be in the loop for continuous dynamic reconfigurations of measurement probes. In the absence of information about past measurements, it may start with an initial set of measurement sessions, possibly randomly seeding a set of starter measurements.

## The Use of P2P Techniques

Several authors (for a discussion on this, see (Samaan & Karmouch, 2009) claim that some level of decentralization plays an important role to perform autonomic actions in a more adequate manner. Different technologies could be employed as an infrastructure of a decentralized ANM system. P2P technology can provide the foundations for increasing the intelligence applied to the control of measurement mechanisms through sophisticated distributed network management algorithms. Network devices could benefit from a network-wide and distributed control of such mechanisms since it is feasible that measurement decisions (e.g., activation of active measurement sessions) are better made by considering the sharing of computational resources and management information. Thus, such decisions may take into account local and remote information as well as consider resources just from the device itself and from remote devices. In addition, the interfaces for the full control of active measurement mechanisms are usually provided only locally on the devices, which also hampers the use of centralized and hierarchical approaches.

The authors advocate for embedding P2P techniques within network devices in order to use autonomic control loops to make decisions about measurement sessions. A pragmatic approach to deploy such techniques in the control of the activation of active measurement sessions is to define principles to guide this deployment (Nobre, & Granville, 2017). Specifically, the authors advocate for: (1) network devices to implement an autonomic function that monitors service levels for violations of SLOs and that determines which measurement sessions to set up at any given time based on current and past observations of the node's service levels and of other peer nodes; (2) autonomic functions to provision the measurement overlay, since the provisioning of the such overlay should be transparent for the network administrator and should facilitate the exchange of data between different nodes to share measurement results so that each node can refine its measurement strategy not just based on its own observations, but also on observations from its peers, and; (3) remote measurements to optimize resource consumption, since they allow nodes to coordinate their measurements to obtain the best possible test coverage. Nodes may utilize observations that are made by their measurement peers in order to conclude which measurement targets may be more critical than others and to ensure that proper overall measurement coverage is obtained. Put simply, the above principles try to capture the common sense used by network administrators when using active measurement mechanisms to detect SLA violations.

The utilization of past service level measurement results can be exploited to detect whether a destination is likely to disrespect SLOs or not. To do so, descriptive statistical metrics can be used to measure how close past service level measurement results are regarding the SLO for a given destination. If the past measurements results for a given destination are close to a SLO, then the probability of activating a measurement session for this destination should increase. This is done by local logic, i.e., an application that runs locally on the network devices. By performing these functions locally and autonomously on the device itself, the measurements to conduct can be modified quickly, based on local observations while taking local resource availability into account. This enables more robust and more reactive solutions so that they can rapidly change service levels compared to central coordination designs. For example, a node could decide to adjust the amount of synthetic test traffic being sent during the measurement

itself, depending on results observed so far on this node, and depending on other concurrent measurement sessions.

A distributed autonomic solution also allows nodes to coordinate their probing decisions to collectively achieve the best possible measurement coverage. Service level measurement results are produced by active measurement mechanisms around the network infrastructure. This information can speed up the detection of SLA violations and improve the number of detected SLA violations. In this context, human administrators can usually predict if SLA violations are likely to happen in a specific region of the network, by using information from measurements of other regions of the network. This is possible because administrators can use their experience and knowledge to infer the relation among the links within the network infrastructure. Because the number of resources available for SLA monitoring is limited, a node may be interested in identifying other nodes whose observations are similar to its own. This helps a node prioritize and decide which other nodes to coordinate and exchange data with.

P2P techniques can be used to capture one of the behaviors commonly adopted by human administrators to detect SLA violations: the sharing of measurement results. For example, if one device detects that a remote destination is about to violate an SLO, other devices may conduct additional measurements to the same destination or other destinations in its proximity. However, it is important to define which network devices are prone to share measurement results. For any given network device, the exchange of data may be more important with some devices than with others. Defining the network devices that exchange measurement data creates a new topology. Different approaches could be used to define this topology (e.g., correlated peers, Nobre, Granville, Clemm, & Prieto, 2012). Using the P2P overlay, measurements can be coordinated among different network devices to avoid hitting the same destination at the same time and to share results that may be useful for future probe placement.

## Implementation Considerations

Network devices have increased substantially their level of programmability. Thus, management software can be embedded to control the activation of active measurement mechanisms. Embedded management peers can have direct access to the internal API of the active measurement mechanism, which could foster the configuration of measurement sessions. Since activating measurement sessions should be a dynamic process in modern network infrastructures, such fostering can improve measurement efficiency (e.g., in terms of monitoring coverage). Considering the ANIMA Autonomic Networking Infrastructure (ANI), Autonomic Service Agents (ASAs) can be implemented on nodes in the network, by using the devices' programmability interfaces (such as Juniper's Junos API or Cisco's Embedded Event Manager). Finally, it should be noted that there are multiple deployment scenarios, including scenarios that involve physical devices virtualized infrastructures hosting autonomic functions.

Embedded P2PBNM systems also make the growth of management resources more "organic" since they can grow without requiring a fork-lift upgrade (i.e., new devices can host by default new management peers). In this context, these systems could grow as new devices are added, bundled with embedded management peers. In order to bootstrap peer selection, each node should use its known neighbors (e.g., based upon the entries maintained by its FIB and RIB tables) as initial seeds to identify possible peers. In addition, an autonomic solution will be useful if topology information and network discovery functions are provided by the underlying ANI. The Autonomic Control Plane defined by the ANIMA ANI (Autonomic Network Infrastructure) provides an ideal candidate for the embedded P2P overlay to run on.

Each device needs to have self-knowledge about the local SLA monitoring. Thus, it is necessary to access and store historical measurement data and SLOs. In addition, measurement data and SLOs can be complemented with passive measurements such as flow data (to identify network destinations that are currently popular and critical). Since interoperability in a heterogeneous network monitoring environment is necessary, only a minimum set of features found on different active measurement mechanisms (e.g., OWAMP, TWAMP, and Cisco's Service Level Assurance Protocol) is required. In addition, the devices would run algorithms that can decide which probes should be activated at any given time. The choice of the algorithm for a specific situation would be made by means of autonomic networking. In this context, nodes would have a repository of algorithms (and correlation functions) that could fit given network conditions.

The autonomic solution for detecting SLA violations should assume that a typical infrastructure will have multiple network segments, ASs, and a reasonably large number of network devices with the capacity of performing measurement sessions. Such solution should also consider that multiple SLOs need to be achieved at any given time. In addition, the autonomic solution should make possible for nodes (or more specifically, the ASAs that are supported by these nodes) to autonomously set up measurement sessions without having to rely on a central management system or controller to perform configuration operations associated with configuring measurement probes and responders.

## Use Case Limitations

Despite the several benefits an autonomic solution for the distributed detection of SLA violations provides, some limitations can be described considering the use case presented in RFC8316. This section describes such limitations.

Full autonomic solutions minimize human intervention in the distributed detection of SLA violations. In this context, practical autonomic features can, at least, minimize such intervention. However, there are some processes that still require a human administrator. The policies regarding how closely to monitor the network for SLA violations and the resource budget that is assigned to network devices for measurement operations may be set by a human administrator. With that budget, the number of SLO violations that are detected can be improved by the autonomic solution.

The use case considers features commonly supported by widely known active measurement mechanisms, such as TWAMP and IPSLA. In this context, the chosen mechanism for SLA monitoring does not need any modification to be controlled by the approach described in this use case. Furthermore, it is assumed that there is an open interface for the activation of measurement sessions in the network devices which support the measurement mechanism themselves.

Security issues are considered orthogonal to the present use case. However, the authors are aware that these use case has security implications and the authors regard options to overcome these implications as future work. In any case, the security of the autonomic detection of SLA violations hinges on the security of the network underlay, which is the Autonomic Control Plane in the case of an ANIMA ANI. If the Autonomic Control Plane were to be compromised, an attacker could undermine the effectiveness of measurement coordination by reporting fraudulent measurement results to peers. This would cause measurement probes to be deployed in an ineffective manner that would increase the likelihood that violations of SLOs go undetected.

## FUTURE RESEARCH DIRECTIONS

The present use case is intended to be an initial step towards autonomic detection of SLA violations. It is also important to investigate how coordination features can enable composite measurement tasks. In addition, refinements in the definition of correlated peers can be included to allow a more selective peering. For example, throttling of overlay traffic can be introduced for "popular" peers. Furthermore, the information about correlated peers can have another usage, e.g., allow inferences about the underlying (physical) topology (i.e., the discovery of the topology of the network substrate).

The security of the autonomic detection of SLA violation hinges on the security of the deployment of such mechanism for autonomic functions and the ANI. In this use case, if the autonomic function that conducts the service-level measurements is hijacked by an attacker, such attacker could try to exhaust or exceed the resources that should be spent on autonomic measurements in order to deplete network resources. This could include network bandwidth due to higher-than-necessary volumes of synthetic test traffic generated by measurement probes. Furthermore, this could also lead to the report of misleading results, thereby resulting in a suboptimal selection of measurement targets. This could increase the likelihood that service-level violations go undetected. Finally, the (ANIMA) ANI could also be attacked, for example through the denial of service regarding the Autonomic Control Plane (ACP) availability and impersonating an autonomic node to participate in the ACP.

Future research and future standards on autonomic networking must consider highly virtualized and programmable infrastructures. In this context, it is necessary to investigate deployment issues in conjunction with Virtual Network Functions (VNFs). In addition, the deployment of new network technologies is typically a time-consuming and labor-intensive task. Thus, SDN techniques may help deploy solutions for an autonomic detection of SLA violations. The NMRG-conducted work on AN with the publication of RFC7575 (Behringer, Bjarnason, Jiang, Carpenter, Pritikin, Ciavaglia, & Clemm, 2015) and RFC7576 (Jiang, Carpenter, & Behringer, 2015) primarily focused on node-level aspects. In this context, the standardization of several ANI required elements and technologies remains an open question since the current standards are limited to specific ANI mechanisms. For example, it is still necessary new initiatives on (Policy) Intent, Use Cases, ASAs, etc., considering the ANI operation.

## CONCLUSION

Violations of SLOs can be associated with significant financial loss in several network infrastructures. Such loss can be divided into two categories. First, there is the loss that can be incurred by the user of a service when the agreed service levels are not met. Second, there is the loss that is incurred by the provider of a service who is unable to achieve contractually defined SLOs. Those losses can take several forms, such as penalties for violating the service level agreement and even loss of future revenue due to a reduced customer satisfaction. Hence, SLOs are a key concern for the service provider. In order to ensure that SLOs are not being violated, service levels need to be continuously monitored in order to know, for example, when mitigation actions need to be taken. To that end, service-level measurements must take place. However, the deployment and operation of these measurements is a demanding task for human administrators.

The problem to be solved by AN in the present use case is how to steer the process of measurement session activation by a solution that sets all necessary parameters for this activation to operate efficiently, reliably, and securely, with no required human intervention other than setting global policies. The authors advocate the use of P2P techniques to increase the potential number of detected SLA violations by means of active measurement mechanisms. This use leads to the introduction of key concepts to support a self-organizing, embedded P2P measurement overlay that uses the capabilities of the network devices to control session activation. In practice, these factors combine to maximize the likelihood of SLA violations being detected while operating within a given resource budget, allowing a continuous measurement strategy. The solution takes into account past measurement results, observations of other measurements such as link utilization or flow data, measurement results shared between network devices, and future measurement activities coordinated among nodes. As a result, the proposed P2P-based solution could decrease the time to detect SLA violations and help reduce the workload of human administrators.

## ACKNOWLEDGMENT

The initial presentation of the Internet-Draft that led to RFC8316 was supported by the Internet Society [Fellowship to the Internet Engineering Task Force (IETF)]. This presentation was made during the IETF Meeting 90 (2014), Toronto, Canada.

## REFERENCES

Agoulmine, N., Balasubramaniam, S., Botvich, D., Strassner, J., Lehtihet, E., & Donnelly, W. (2006). Challenges for autonomic network management. In *Proceedings of the IEEE International Workshop on Modeling Autonomic Communications Environment (MACE*). IEEE

Behringer, M., Bjarnason, S., Jiang, S., Carpenter, B., Pritikin, M., Ciavaglia, L., & Clemm, A. (2015). *Autonomic networking: Definitions and design goals. (RFC7575)*. Internet Engineering Task Force. doi:10.17487/RFC7575

Bouabene, G., Jelger, C., Tschudin, C., Schmid, S., Keller, A., & May, M. (2010). The autonomic network architecture (ANA). *IEEE Journal on Selected Areas in Communications*, *28*(1), 4–14. doi:10.1109/JSAC.2010.100102

Burger, E. W., & Seedorf, J. (2009). *Application-layer traffic optimization (ALTO) problem statement (RFC5693)*. Internet Engineering Task Force.

Chaparadza, R., Papavassiliou, S., Kastrinogiannis, T., Vigoureux, M., Dotaro, E., Davy, A., & Wilson, M. (2009). *Creating a viable Evolution Path towards Self-Managing Future Internet via a Standardizable Reference Model for Autonomic Network Engineering*. In Future Internet Assembly.

Chen, M., Castaldelli, L., Mirsky, G., Mizrahi, T., Fioccola, G., Capello, A., & Cociglio, M. (2018). *Alternate-Marking Method for Passive and Hybrid Performance Monitoring (RFC8321)*. Internet Engineering Task Force.

Chiba, M., Clemm, A., Medley, S., Salowey, J., Thombare, S., & Yedavalli, E. (2013). *Cisco service-level assurance protocol (RFC6812)*. Internet Engineering Task Force.

Claise, B. (2008). *Specification of the IP flow information export (IPFIX) protocol for the exchange of IP traffic flow information (RFC5101)*. Internet Engineering Task Force.

Claise, B., Trammell, B., & Aitken, P. (2013). *Specification of the IP flow information export (IPFIX) protocol for the exchange of flow information (RFC7011)*. Internet Engineering Task Force.

Di Pietro, A., Huici, F., Costantini, D., & Niccolini, S. (2010). *Decon: Decentralized coordination for large-scale flow monitoring*. In *INFOCOM IEEE Conference on Computer Communications Workshops*, 2010 (pp. 1-5). IEEE. 10.1109/INFCOMW.2010.5466642

Duffield, N., Chiou, D., Claise, B., Greenberg, A., Grossglauser, M., & Rexford, J. (2009). *A framework for packet selection and reporting (RFC5474)*. Internet Engineering Task Force.

Elkins, N., & Hamilton, R. (2017). *IPv6 Performance and Diagnostic Metrics (PDM) Destination Option (RFC8250)*. Internet Engineering Task Force. doi:10.17487/RFC8250

Fiorese, A., Simões, P., & Boavida, F. (2009). A P2P-based approach to cross-domain network and service management. In *IFIP International Conference on Autonomous Infrastructure, Management and Security* (pp. 179-182). Springer.

Hedayat, K., Krzanowski, R., Morton, A., Yum, K., & Babiarz, J. (2008). *A two-way active measurement protocol (TWAMP) (RFC5357)*. Internet Engineering Task Force.

Jiang, S., Carpenter, B., & Behringer, M. (2015). *General gap analysis for autonomic networking (RFC7576)*. Internet Engineering Task Force. doi:10.17487/RFC7576

Jones, S. (2005). Toward an acceptable definition of service service-oriented architecture. *IEEE Software. IEEE.*, *22*(3), 87–93. doi:10.1109/MS.2005.80

Mortier, R., & Kiciman, E. (2006, September). Autonomic network management: some pragmatic considerations. In *Proceedings of the 2006 SIGCOMM workshop on Internet network management* (pp. 89-93). ACM. 10.1145/1162638.1162653

Morton, A. (2016). *Active and Passive Metrics and Methods (with Hybrid Types In-Between) (RFC7799)*. Internet Engineering Task Force. doi:10.17487/RFC7799

Nobre, J., Granville, L., Clemm, A., & Prieto, A. G. (2018). *Autonomic Networking Use Case for Distributed Detection of Service Level Agreement (SLA) Violations (RFC8316)*. Internet Engineering Task Force. doi:10.17487/RFC8316

Nobre, J. C., & Granville, L. Z. (2017, May). Decentralized detection of violations of service level agreements using peer-to-peer technology. In *Integrated Network and Service Management (IM), 2017 IFIP/IEEE Symposium on* (pp. 835-840). IEEE.

Nobre, J. C., Granville, L. Z., Clemm, A., & Prieto, A. G. (2012). Decentralized detection of SLA violations using P2P technology. In *Proceedings of the 8th International Conference on Network and Service Management* (pp. 100-107). International Federation for Information Processing.

Nobre, J. C., Mozzaquatro, B. A., & Granville, L. Z. (2018). Network-Wide Initiatives to Control Measurement Mechanisms: A Survey. *IEEE Communications Surveys and Tutorials*, *20*(2), 1475–1491. doi:10.1109/COMST.2018.2797170

Pras, A., Schonwalder, J., Burgess, M., Festor, O., Perez, G. M., Stadler, R., & Stiller, B. (2007). Key research challenges in network management. *IEEE Communications Magazine*, *45*(10), 104–110. doi:10.1109/MCOM.2007.4342832

Samaan, N., & Karmouch, A. (2009). Towards autonomic network management: An analysis of current and future research directions. *IEEE Communications Surveys and Tutorials*, *11*(3), 22–36. doi:10.1109/SURV.2009.090303

Sekar, V., Reiter, M. K., Willinger, W., Zhang, H., Kompella, R. R., & Andersen, D. G. (2008). cSamp: A System for Network-Wide Flow Monitoring. In NSDI (Vol. 8, pp. 233-246). Academic Press.

Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., & Zekauskas, M. (2006). *A One-way Active Measurement Protocol (OWAMP) (RFC4656)*. Internet Engineering Task Force.

Strassner, J. (2004, April). Autonomic networking-theory and practice. In *Network Operations and Management Symposium, 2004. NOMS 2004. IEEE/IFIP* (Vol. 1, pp. 927-Vol). IEEE. 10.1109/NOMS.2004.1317811

Voulgaris, S., Gavidia, D., & Van Steen, M. (2005). Cyclon: Inexpensive membership management for unstructured P2P overlays. *Journal of Network and Systems Management*, *13*(2), 197–217. doi:10.100710922-005-4441-x

## ADDITIONAL READING

Bajpai, V., & Schönwälder, J. (2015). A survey on internet performance measurement platforms and related standardization efforts. *IEEE Communications Surveys and Tutorials*, *17*(3), 1313–1341. doi:10.1109/COMST.2015.2418435

Boucadair, M., Jacquenet, C., & Wang, N. (2014). *IP Connectivity Provisioning Profile (CPP) (RFC7297)*. Internet Engineering Task Force.

Dobson, S., Denazis, S., Fernández, A., Gaïti, D., Gelenbe, E., Massacci, F., ... Zambonelli, F. (2006). A survey of autonomic communications. *ACM Transactions on Autonomous and Adaptive Systems*, *1*(2), 223–259. doi:10.1145/1186778.1186782

dos Santos, C. R. P., Famaey, J., Schönwälder, J., Granville, L. Z., Pras, A., & De Turck, F. (2016). Taxonomy for the network and service management research field. *Journal of Network and Systems Management*, *24*(3), 764–787. doi:10.100710922-015-9363-7

Dreibholz, T., Rathgeb, E. P., Rungeler, I., Seggelmann, R., Tuxen, M., & Stewart, R. R. (2011). Stream control transmission protocol: Past, current, and future standardization activities. *IEEE Communications Magazine*, *49*(4), 82–88. doi:10.1109/MCOM.2011.5741151

Dressler, F., & Akan, O. B. (2010). A survey on bio-inspired networking. *Computer Networks*, *54*(6), 881–900. doi:10.1016/j.comnet.2009.10.024

Eardley, P., Morton, A., Bagnulo, M., Burbridge, T., Aitken, P., & Akhter, A. (2015). *A framework for Large-scale Measurement of Broadband Performance (LMAP) (RFC7594)*. Internet Engineering Task Force. doi:10.17487/RFC7594

Nobre, J. C. (2016). *Decentralized detection of violations of service level agreements using peer-to-peer technology*. Ph.D Thesis

Strassner, J., Agoulmine, N., & Lehtihet, E. (2006, April). *FOCALE–a novel autonomic computing architecture*. In *Proceedings of the 2006 Latin–American Autonomic Computing Symposium*.

## KEY TERMS AND DEFINITIONS

**Active Measurements:** Techniques to measure service levels that involve generating and observing synthetic test traffic.

**Autonomic Network:** A network containing exclusively autonomic nodes, requiring no configuration, and deriving all required information through self-knowledge, discovery, or intent.

**Autonomic Service Agent (ASA):** An agent implemented on an autonomic node that implements an autonomic function, either in part (in the case of a distributed function, as in the context of this chapter) or whole.

**Measurement Session:** A communications association between a probe and a responder used to send and reflect synthetic test traffic for active measurements.

**P2P:** Peer-to-peer.

**Passive Measurements:** Techniques used to measure service levels based on observation of production traffic.

**Probe:** The source of synthetic test traffic in an active measurement.

**Responder:** The destination for synthetic test traffic in an active measurement.

**SLA:** Service level agreement.

**SLO:** Service level objective.