Industry-Academia Research towards Future Network Intelligence: The NG-CDI Prosperity Partnership

Journal:	*IEEE Network Magazine*
Manuscript ID	NETWORK-21-00405.R2
Topic or Series:	January 2022/Bridging the Gap between Industry and Academia for Networking Research
Date Submitted by the Author:	19-Nov-2021
Complete List of Authors:	Race, Nicholas; Lancaster University, School of Computing & Communications Eckley, Idris; Lancaster University, Mathematics and Statistics Parlikad, Ajith; University of Cambridge Institute for Manufacturing Rotsos, Charalampos; Lancaster University, School of Computing and Communications Wang, Ning; University of Surrey Piechocki, Robert; University of Bristol, School of Computer Science, Electrical and Electronic Eng, and Eng Maths Stiles, Philip; University of Cambridge Judge Business School Parekh, Arjun; British Telecom Burbridge, Trevor; British Telecom, Applied Research Willis, Peter; BT, R&I Cassidy, Stephen; British Telecom
Key Words:	industry, academia, partnership, autonomic, future networks

Industry-Academia Research towards Future Network Intelligence: The NG-CDI Prosperity Partnership

1 Abstract

Ever since the first automation provided by the introduction of the Strowger telephone exchange in the late 19th century, networks have been increasingly automated. Fast forward to 2022, and the challenge facing network providers is scaling up this level of automation considering massive increases in complexity, new levels of agility to operate services, and rising demand from customers within the modern telecommunications ecosystem. This paper describes a significant new industry-academia partnership to address these challenges: Next Generation Converged Digital Infrastructure (NG-CDI) is creating a vision for the building and operation of a future-proof network infrastructure and its autonomic management. In this paper we highlight three exemplar activities within the NG-CDI research programme that illustrate the benefits of taking a highly collaborative interdisciplinary approach and show how academia and industry working closely together has delivered a range of direct and positive impacts on the business.

2 Introduction

Network providers are predicting a huge growth in new services and applications, with tens of billions of devices, sensors, vehicles and people to become interconnected over the next 10 to 15 years. All of this will place unprecedented demand on the underlying infrastructure. Traditionally, the deployment of new services has involved reinvestment in infrastructure, extensive pre-testing, and people-intensive service support in operation, requiring several hundred people to deliver. However, future services will change ever more rapidly – and unpredictably – and therefore organizations need to drastically reduce the time it takes for new services to be developed, trialled and launched. This requires radical improvements in the agility and responsiveness of the network infrastructure itself: the infrastructure will need to support near real-time monitoring of performance, self-diagnose problem states and enable rapid and informed human intervention when needed. The ambition is to reduce service costs as well as providing a framework to spin out innovation in days rather than years. The core infrastructure will not only provide huge capital and operational cost savings, significant in itself, but also greater growth potential, since the cost of innovation and experimentation will be reduced, and its speed increased.

To address these challenges a new industry-academic partnership, Next Generation Converged Digital Infrastructure (NG-CDI), is creating a vision for the building and operation of a future-proof network infrastructure and its autonomic management. Such an infrastructure must be capable of fast and efficient service innovation and co-creation with a wide variety of customers. NG-CDI promises completely new ways of operating the infrastructure. Recent advances in programmable network interfaces, and model-driven networking provide the possibility of closed-loop, self- optimizing, and self-healing operations. NG-CDI builds on these breakthroughs to deliver greater economies and customer value. Operating changes of this scale need not only radical technological solutions, but also changes to the organisation itself.

This paper highlights the benefit of taking a collaborate academia-industry approach to addressing future networking challenges based on real-world problems. The paper starts by describing the business challenges that led to the formation of the NG-CDI industry-academia partnership, and then outlines the approach taken to conducting research between the stakeholders. This is followed by an overview of the NG-CDI research programme and architecture. We then describe in detail three research exemplar activities; each of which is at a different stage of maturity, but nevertheless each is having an impact within the industry at a different stage of the transformation journey. Finally, we conclude the paper and identify a number of areas future work based around the risk and governance implications associated with the introduction of the NG-CDI architecture.

3 Business Challenges and Approach to Research

NG-CDI is an ambitious collaboration, established between British Telecom (BT) and four of BT's long-term strategic University partners, and supported by the UK's Engineering and Physics Sciences Research Council 'Prosperity Partnership' scheme. The Prosperity Partnership scheme is designed to foster strategic, research-based partnerships between industry and academia through co-investment in shared research challenges.

NG-CDI is directly addressing the challenges associated with the huge expansion in the scale and value of modern networks. In the past, routine human actions have been scaled up through automation. This has seen human effort transferred to engineering design, proposition development and other higher-level activities. The challenge now facing networks is to extend these activities to manage the massive increase in complexity, speed of change, and customer responsiveness demanded by the modern telecommunications ecosystem [1].

Addressing these challenges requires transformational research to broaden the understanding and knowledge base, not only of the capabilities underpinning any digital infrastructure but also the ability to manage and operate such a complex system.

The fundamental approach taken by NG-CDI was to bring together a multi-disciplinary team of academic and industry researchers to co-create the research programme and evolve it during the project. This enables the research agenda to develop as new discoveries were made and new business drivers identified. The BT researchers articulated challenges and use cases derived from co-working with the business areas of the company, provided masses of business and operational data for the development of models and algorithms, and added their own expertise and experiences. The academics, each pre-eminent in their field, were able to open up the range of techniques and approaches that could be applied, using research from other fields, or by developing new advanced methodologies. This approach maximises the value of the project by mixing exploratory high potential research from the universities with the problem-focused research and exploitation opportunity from the industrial researchers.

Given the wide scope and potential impact, it was important to identify and work with a wide range of relevant BT stakeholders, including research and business areas. Coherent visions of the value and potential impact of the work were generated to gain traction and co-operation. These interactions helped articulate concrete use cases to focus the work. Close co-operation in creating chosen "quick wins" using data provided by BT stakeholders helped demonstrate tangible benefits and build the enthusiasm to work with academics. Broader support and interest was achieved through a set of talks from the senior academics to company-wide audiences as part of BT's Thought Leadership programme, each with around 100 attendees representing all the lines of business. At a more operational level, weekly meetings between BT personnel and the research team generated a sense of pace across the project through feedback and direction. Sub-groups would meet regularly to advance specific topics, and these interactions have aligned perspectives and established common languages across the different disciplines involved. This approach has been instrumental in cementing the overall vision, supported by the underlying detail.

4 NG-CDI Research Overview

Realizing the NGI-CDI vision required an intimate understanding of operating national and international digital infrastructure, and world-class expertise in the areas of data analytics, machine learning, cyber-physical systems, network functions virtualization, networked systems, asset management and business innovation.

Using a combination of approaches, enables improved efficiency and a more direct focus on customer and commercial benefits, combining intelligent infrastructure and autonomic control, based on customer and business targets. Delivering "intents" rather than low-level technological specifications improves the level of automation. An Intent-Based Networking (IBN) approach designates a high-level requirement that can be expressed from an external client, application, or owned by the network operator. Once an intent arrives into the system, it passes through different stages of translation before reaching the management plane. This process converts a high-level expression to something more technical and feasible within the management plane before being configured/enforced at the relevant network devices.

Iterating the requirements through simulations results in machine-readable intents (formalised requirements) which can be used automatically to orchestrate software-based network and service functions, and deployed into the infrastructure to deliver the service. The service capacity can be scaled up or down without needing to install dedicated equipment. The agreed service levels will need to be managed dynamically. In the real world there will be disturbances such as surges in traffic demand, equipment failures, data errors, engineering works. Some of these events will be manageable by self-learning software agent control, which in real time can find the best available new balance between the various requirements – such as maintaining certain service levels subject to cost constraints. The control algorithms use network events: traffic, telemetry etc., to learn about problem states and remaining useful life of network elements. This updates the agents accordingly to self-optimise the service-level intents. For this autonomous control, it is crucially important that we balance a new category of costs/risks. A learning algorithm needs to know what accuracy of decision is required, how fast the learning rates need to be, the required speeds of response – all of which are fundamentally related. This involves new ways of judging and making decisions on the extent of prognostic maintenance used to pre-empt service issues, based on the likely scale of consequences. NG-CDI is looking at model-supported business decision-making processes and cultures that the industry will increasingly need.

Autonomic capability will be distributed through different domains. For example, many aspects of a 5G network will be self-optimising. Centralised control will not scale and to support the key business functions we need a sufficient representation of the knowledge needed. This will be a mix of types. Some will be pre-stored scenarios built from modelling or machine-learning – which can be enacted when appropriate. Some will be selected or aggregated data and used in real-time monitoring, tuned to the right response rate and accuracy. Real-time data streams will be used to diagnose network issues, or respond as necessary. This requires sophisticated new statistical techniques for detecting anomalies in massive real-time data flows, distinguishing them from normal statistics.

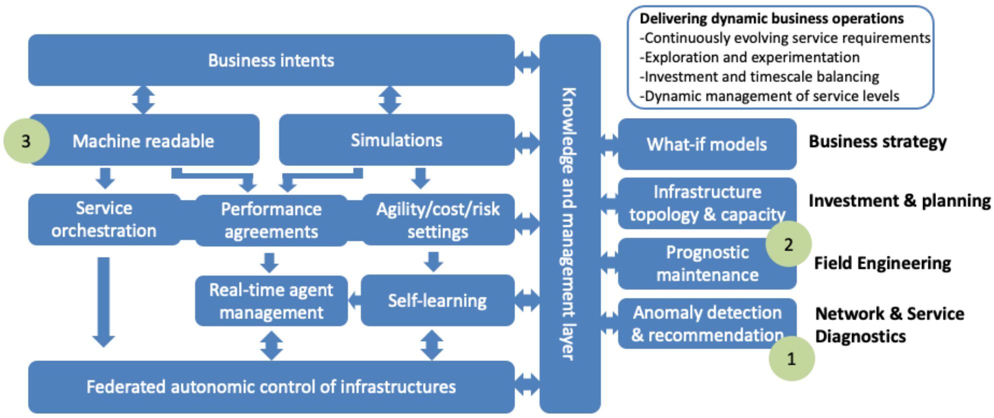

Figure 1 - NG-CDI Architecture

The project is supported by an underlying architecture (Figure 1). The figure highlights three exemplars of activity within the overall NG-CDI programme that are described in the remainder of the paper: 1) Anomaly Detection, 2) Prognostic Maintenance and 3) Business Intents. These three exemplars encompass representative technical challenges faced by BT towards future evolved network autonomics in different network management applications.

5 NG-CDI Exemplar Activities

Delivery of the full benefits of the project depends on building a range of outputs, encompassing different areas of impact and on different timescales. For example, a technological and business long-term vision needs to be developed across the industry to develop paths towards manufacture and implementation based on international scale economics. Alongside this long-term influence, immediate opportunities have been created, delivering benefits in the shape of smarter customer processes, and improved service levels. Medium term opportunities are developed through building conversations across the business. These are often stimulated by proof-of-concept demonstrations which show the art of the possible through real or simulated networks and data.

Close co-working between university and BT researchers has enabled benefits across this range of impacts. Techniques have been developed based on massive operational data from BT's network and processes. This means the solutions are bedded in reality and more readily assimilated by the business. BT researchers provide a bridge to relevant business and operational areas, translating their research knowledge to articulate the benefits and impacts in the language of relevant business domains. The existence of network testbeds provides further opportunities by enabling the deployment trial services in real-world environments, such as 5G O-RAN, from which operational and user experience can be tested.

In this section we highlight examples of successful impact areas, which serve to illustrate the different types of business impact and their relationship to the ambitions for the converged digital infrastructure. Devolving processes to autonomic control increases the responsivity to the dynamic network environment and changing requirements. It releases human effort to concentrate on higher value activities.

NG-CDI has addressed these opportunities in the following topic areas. *Anomaly Detection* enables the business to discern whether a pattern of network events is one that the autonomic system needs to respond to, and if so without human intervention, or whether human expertise is needed. Systems which learn from network and service data can be used optimise business decisions and processes such as *Prognostic Maintenance*. This enables the business to continually balance between reactive and proactive maintenance to optimise the economics of service delivery. Enabling continuous change in this way can be extended to the customer through *Business Intents*. This enables a more direct translation between the language of business needs of balancing organisational risks, costs and performance to the technical 'instructions' necessary for effective orchestration of the service and the supporting autonomic processes. All the topics described play important roles in reducing the commercial risks inherent in automated systems as well as ensuring safe operation. The integration of these aspects to provide a rich interface to the key business decision processes is an active area of research.

1) Anomaly Detection

Within the current telecoms network environment, we face many data streams that need to be carefully monitored to help ensure the successful performance of the infrastructure. The collection of these streams is crucial to observing and understanding the behaviour of a network that is driven by the behaviour of millions of users and applications. However, the scale of even current-day networks, in terms of the equipment and the number of metrics that can be monitored, means that it is no longer possible to rely on expert users visually reviewing data streams, except at the macro-scale where many smaller problems may be missed. As we look to the data-driven, autonomic networks of the future, the volume and variety of these streams will increase.

Nowhere is this more apparent than in the monitoring of network operational data streams. Here the aim is to rapidly, and accurately, analyse a stream as it is observed to identify those anomalous periods that may be indicative of operational challenges. Timely and accurate detection of such anomalies is critical to help minimise operational disruption and ensure the smooth running of the network. In essence, the problem we need to consider is how to identify anomalous periods efficiently and accurately from the baseline of everyday performance. Consequently, some of the simplest data analyses we might undertake consist of identifying in real-time whether anomalies have occurred and whether such anomalies are point, i.e., a single outlying observation, or collective anomalies. Figure 2 shows an example of this for a sample of network throughput data.

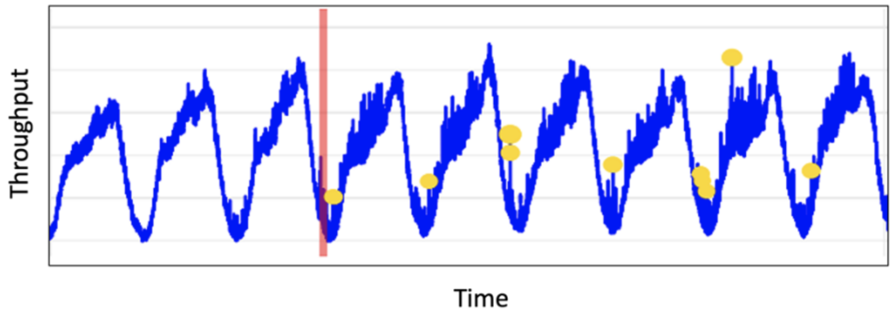

Figure 2 – Analysis of sample network throughput data, identifying periods of typical and anomalous throughput. Note that both point (yellow dots) and collective (red band) anomalies are identified.

We have developed a new suite of computationally efficient anomaly detection methods including [4-5]. Built upon the rigorous foundations of statistical inference, the resulting anomaly methods can be run in real-time to identify whether the raw data might, for example, be a fleeting point anomaly or a more persistent collective anomaly from the typical `baseline'. Within this suite, the sequential collective and point anomaly (SCAPA) approach [4], provides a highly effective mechanism to detect anomalies within univariate data sequences.

SCAPA is based on a dynamic programming algorithm, analysing each new data point as it is received. Broadly speaking, the approach works as follows: Firstly, SCAPA assumes that all the data is drawn from the same distribution and assigns a cost to this. The methods then seeks to segment the data, to reduce this cost. Individual points are also removed in order to further reduce the cost. When this occurs, the point is flagged as a point anomaly.

One particular novelty of SCAPA, as a method, is that it re-assesses these anomalies each time a new observation is received. In addition to discerning the nature of the change, the theory developed also provides understanding of key questions such as (i) when is it appropriate to use this approach; (ii) the likely delay between anomaly occurrence and detection and (iii) the amount of data needed before an anomaly is randomly observed.

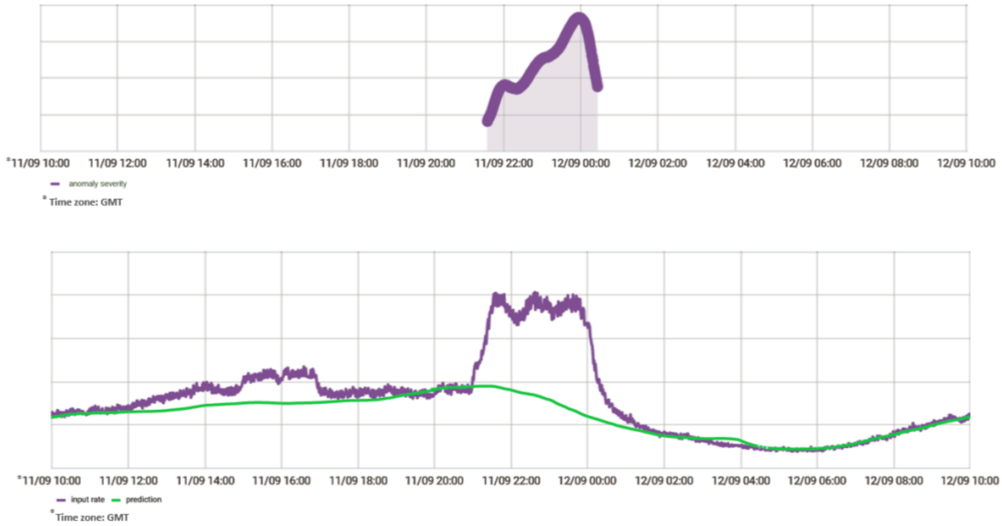

Figure 3 – Sequential CAPA in action within the BT network operations dashboard during the US Open tennis championship final

Through the programme's close collaboration between the academic team and BT's researchers at Martlesham the anomaly methods developed in the programme have been embedded in a number of different operational areas, including BT's Internet Peering Platform. Here, the anomaly tools [3] developed in [4] are used to monitor the platform in real time, triggering anomalies to BT's network operations teams to help them monitor and assure the performance of this critical digital infrastructure that connects millions of users with other network and content providers.

This can be seen in Figure 3, where the operator has drilled down from aggregate views of network anomalies to focus on the data rate through a single network interface. The top part of the dashboard shows how SCAPA is reacting to the emerging anomaly, increasing severity and duration of the collective anomaly with both the rate of increase of the data rate and the duration of the event. The lower part of the dashboard shows the data rate telemetry and the predicted median based upon historical data. This is used in conjunction with the predicted standard deviation of the residuals to normalise the data rate telemetry before presentation to SCAPA. We can see that these steps form a pipeline of processing with both batch (recalculating the predicted median and standard deviation of residuals) and streaming (normalisation of the telemetry and anomaly calculation) operations. In the current implementation the streaming operations are performed using Apache Beam with side-inputs to receive periodic updates to the predictions.

BT is actively working on applying SCAPA and other anomaly detection algorithms to further network and service platforms including the UK core network, Fixed and Radio Access Networks, Content Delivery Networks and IP Voice services.

2) Prognostic Maintenance

Reliability of network equipment is critical to maintaining high levels of service assurance to customers. One of the key benefits of digitalization is the ability to monitor critical metrics that helps understand the health of equipment and to use the data for their effective management and maintenance. In this context, the ability to move from reactive to predictive maintenance practices have become popular across different industry sectors, driven by emerging data-centric approaches for failure diagnosis and prognosis. In particular, the use of advanced machine learning algorithms has improved our ability to exploit monitoring data and predict equipment failures.

However, the main problem with data-driven prognostics is that they rely on large amounts of historical failure data to estimate model parameters effectively. The availability of historical failure data is limited due to two major reasons: (i) over-protective maintenance and replacement regimes; and (ii) highly reliable equipment. This causes failures to be rare, and leads to the problem of limited failure data availability for data-driven prognostics of network equipment, which causes prognostics predictions to be associated with high uncertainty. This was identified by BT network operations as a significant limitation in their ability to determine which event alarms should have priority attention. The manufacturers' equipment alarms include, effectively, very high volumes of false positives, and determining which need to be acted upon is a skilled and time-consuming activity. The operations centre provided high-volume telemetry data from network nodes to enable the research.

In order to address this issue, we developed a technique for generating failure data that realistically reflect the behaviour of degrading equipment (i.e., real-valued failure data) for prognostics under the conditions of limited failure data availability. It allows training datasets used for data-driven prognostics to be augmented so that an increased number of failure data samples is available for prognostics modelling. The methodology generates real-valued failure data using a conditional generative adversarial network (CGAN) by controlling and directing the failure data generation process using auxiliary information pertaining to the failure mode that needs predicting. More specifically, the noise being added to the newly generated failure data samples is conditioned on auxiliary information to prevent different modes of data being generated. Auxiliary information is additional information that adds value to the understanding of failure dynamics of the equipment of interest (e.g., equipment similarity information, expert knowledge on failure causes and failure modes and quality of equipment use).

We applied this technique for predicting the Time-To-Failure (TTF) of telecommunications broadband lines under the conditions of limited failure data availability. To this end, we used the methodology to use expert knowledge on VDSL and ADSL broadband line failure causes (e.g., water ingress into electrical junctions, joints and DPs) to generate real-valued broadband line failure data. Performance was assessed in terms of an "F-score". This is a useful measure in practical decision systems, as it represents the mean of two metrics: precision and recall. Precision is the proportion of true positive predictions amongst all positive (e.g. failures) predictions, and recall is the proportion of true positive predictions with respect to actual positives. Therefore the F-score can be considered as the "risk" of believing in the algorithm as it determines impacts such as customer service levels and cost: i.e. the proportion of real faults intercepted and the wasted effort in responding to false positives. The prognostics performance obtained when prognostics models are trained on the augmented training datasets and evaluated on the test datasets showed significant improvement – an increase of 25% in the F-score for ADSL lines and 13% for VDSL lines compared to the best available existing techniques (Figures 4 and 5).

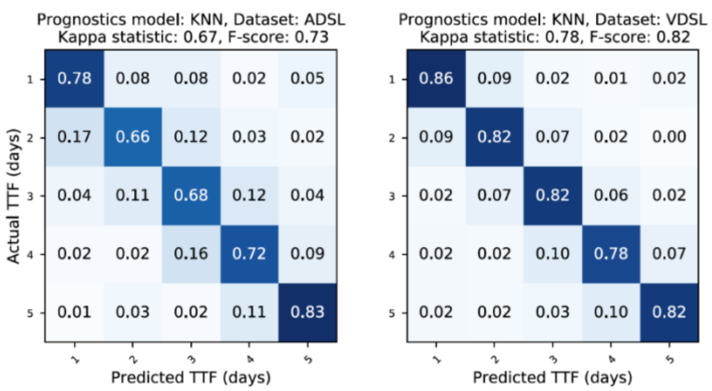

Figure 4: Prognostics performance using best available current technique

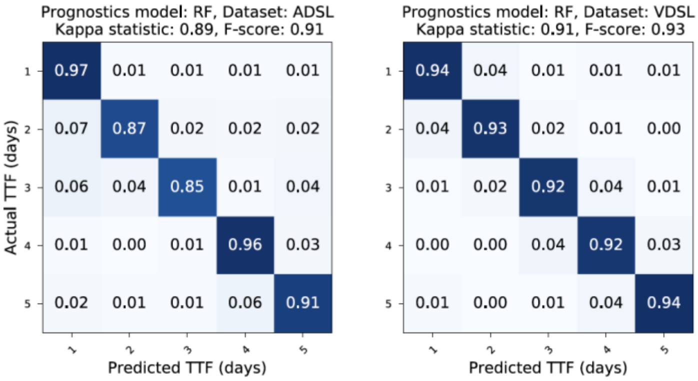

Figure 5: Prognostics performance using our methodology

3) Business Intents

To ensure the adoption of our architecture in production, we opted from the start of the project to use a flexible intent-based approach to expose autonomic control to users and customers. During the early design stages, we conducted a series of interviews with key BT stakeholders (e.g. BT Global Services, Strategic Planning) to better understand business requirements for our architecture and to collect automation use-cases. In our initial requirements analysis, we identified two key design goals: the need for automation in internal business processes and the ability for non-technical users to interact with the autonomic framework. Although several standardized intent models exist, they focus on customer connectivity-based intents and the automatic translation of those intents into appropriate device configurations. Nonetheless, connectivity services depend on "human-centric" business processes, currently unsupported by existing intent models (i.e., billing, logistics). Finally, existing model-based intent systems require high precision in policy expression, which frequently confuses non-technical users, who rely on network managers to eliminate ambiguity when specifying network requirements.

To meet our first design goal, we developed a new intent system which supports control for new resource domains (e.g. orchestration) and provides new types of intents that abstract business processes, including equipment upgrades and service protection. Figure 6 presents an example execution of an equipment upgrade intent [2] as a representative use-case. The use case was developed in collaboration with BT Global Services and allows the network administrator to automatically replace network equipment with a short TTF, estimated by the prognostic maintenance model. Existing equipment upgrade processes typically rely on manual configuration and testing of devices by network administrators, experienced with vendor-specific configuration interfaces. The process usually is repetitive, and human supervision essentially ensures the timely detection of sporadic hardware failures and misconfigurations. Nonetheless, critical upgrades can take several days due to the inability to extensively parallelize the upgrade process, due to limited human resources. In parallel, equipment upgrades lead to unnecessarily long service downtimes to accommodate potential delivery and installation delays.

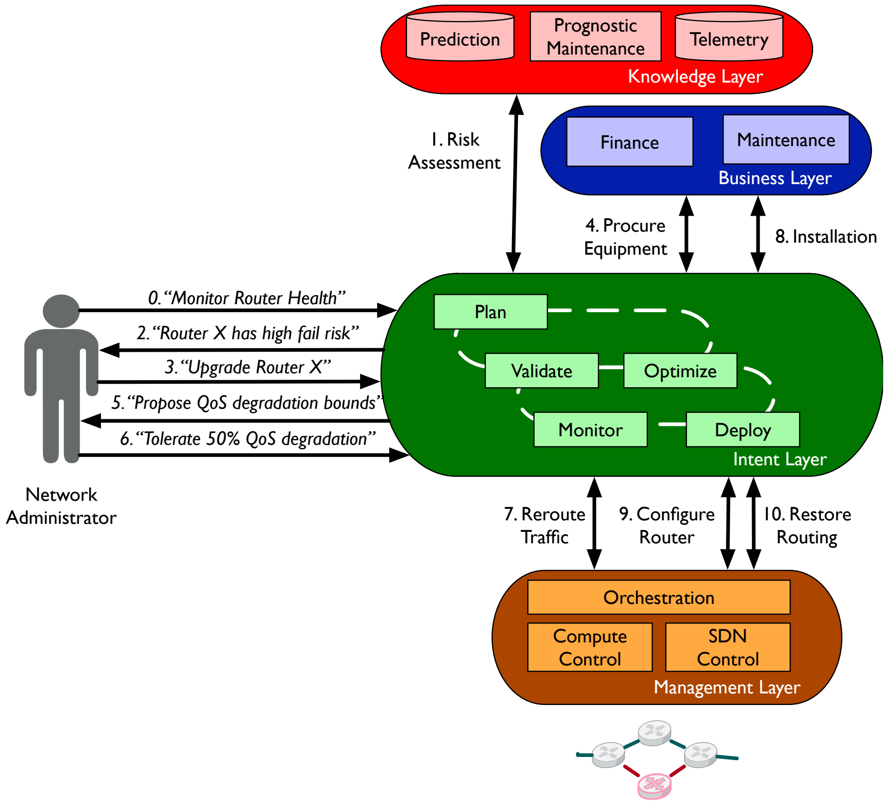

Figure 6: Equipment Health monitor and upgrade intents

The intent uses an Operations Support System (OSS) driver to procure the new equipment, and to schedule infrastructure engineers site visits, once delivery dates are confirmed. In parallel the intent layer can schedule a re-routing operation to the management layer, in order to drain traffic a-priori from the failing equipment. The re-routing process can result in QoS degradation, due to a temporary lack of network resources or further stretched rerouting paths, and the intent layer can negotiate with the network manager possible QoS relaxation windows. Finally, infrastructure engineers can signal the start and end time of the upgrade process to the intent layer, to minimize downtimes, while SDN interfaces can be used for automatic device configuration and fault checking. It is worth highlighting that the integration of the intent and knowledge plane allows risk management strategies specification during the intent planning phase. Specifically, risk metrics can quantify the impact of possible actions (e.g. quantify the real service degradation based on previous traffic patterns) and the intent will select execution plans that reduce operational risks (e.g. schedule upgrades during quiet periods). In parallel, the intent model allows network administrators to associate automatic risk management processes with risk metric thresholds to automate failure response (e.g. if the equipment configuration action fails, the intent must automatically inform an engineer).

To improve the accessibility of our intent system we developed a new prototype interface which helps users to reduce the level of ambiguity during the process of capturing intents. Specifically, we implemented a conversational interface, that uses natural language techniques to analyse and extract user goals. Our prototype interface implementation allows users to interact with our intent layer via popular services, such as Google Assistant and Slack [6]. The interface system explains interactively the impact of an expressed intent, thus allowing users to correct mistakes and reduce ambiguity. In parallel, users can inspect the intent state and receive error notifications throughout the lifecycle of an intent. Our approach enables intent systems to put the user in the loop of the intent process and, in the future, we aim to explore how user input can improve intent validation and re-planning. For example, users can suggest ways to resolve intent conflicts, or provide tiebreakers to the intent optimization process.

6 Conclusions and Future Work

We believe that NG-CDI is a perfect example of effective collaboration between academia and industry in an effort to drive a future networking research agenda based around real-world problems. On the one hand, the osmosis of academic research with network business processes allows the project to identify novel challenges and use-cases for research. On the other hand, academic partners are able to access operational data and even explore opportunities for deployment of models and system in production environments.

The areas of work highlighted in the paper illustrate the benefits of this highly collaborative interdisciplinary approach. The business impacts range from specific applications of ground-breaking new techniques being trialled in volume business processes; through demonstrated visions of how smarter business and network processes could bring customer and business benefit. In addition, the consortial approach has brought thought leadership to help develop strategies in the business as a whole, and the telecoms industry beyond. In each area of work the ongoing collaboration between academia and industry is resulting in a direct and positive impact on the business, as well as benefitting the research community.

We are acutely aware that whilst the introduction of an automated, knowledge-based management architecture promises huge potential for greater and more timely control over our environment, the introduction of such an architecture also introduces new risks. As the technology becomes more complex, we need to avoid it becoming less transparent and accountable.

As part of our Future Work, we are examining the risk and governance implications associated with the introduction of such an architecture. Firstly, we are examining the risks within the architecture itself, covering the development of technologies from design to deployment, testing, re-calibration and revision, and to map the risks involved at each stage. Our approach is informed by research in model risk management. A second aspect of risk concerns governance. Ensuring the elements of the architecture are managed for risk is highly important, and so too is the integration of this architecture into the wider systems and organisation of the business as a whole.

Through working with business and risk management areas of the company we are developing the vision and new approaches that aim not only to maintain robust operations, but also take advantage of the technology to reduce the existing business risks associated with legacy technologies. The introduction of the architecture has strong potential to deliver the future strategy for managing future digital infrastructures. Our integration of model risk management and enterprise risk management will provide a new and integrated framework to assess the introduction of advanced technology and the balance of risks it entails.

7 Acknowledgment

The authors gratefully acknowledge the support of the Next Generation Converged Digital Infrastructure (NG-CDI) Prosperity Partnership project funded by UK's EPSRC and British Telecom plc.

References

[1] M. Bezahaf, D. Hutchison, D. King and N. Race. Internet Evolution: Critical Issues. *Proceedings of the IEEE*, 24(4):5–14, Jul 2020.

[2] M. Bezahaf, E. Davies, C. Rotsos and N. Race, To All Intents and Purposes: Towards Flexible Intent Expression, 1st International Workshop on Intent-based Networking. IEEE, Tokyo, Japan, 2021.

[3] ATM Fisch, D Grose, L Bardwell, IA Eckley and P. Fearnhead (2021). anomaly: Detecting Anomalies in Data. R package version 4.0.1, available from https://CRAN.R-

project.org/package=anomaly

[4] ATM Fisch, L Bardwell, IA Eckley (2020), Real Time Anomaly Detection And Categorisation, arXiv preprint arxiv:2009.06670.

[5] ATM Fisch, IA Eckley, P. Fearnhead (2021), Subset multivariate collective and point anomaly detection, Journal of Computational and Graphical Statistics (to appear).

[6] M. Bezahaf et al., "Self-Generated Intent-Based System," 10th International Conference on Networks of the Future (NoF), 2019, pp. 138-140, doi: 10.1109/NoF47743.2019.9015045.