The Journey Towards Zero 'Preventable Downtime'

TELSTRA

Executive Summary

In this report, we introduce an innovative risk management framework called the **Journey Towards Zero 'Preventable Downtime'** and explore the value that such an approach may bring.

Some of the key insights in this report include:

Measured Network availability performance has plateaued between 99.00%-99.99%[1] which doesn't meet the exponentially increasing demands of the digital economy.

The framework presented in this paper outlines a risk management framework for IT failure modes across the ideate, create, release and operate phases of the life cycle. At each stage, a failure mode is allocated a risk ranking derived from the severity, occurrence and ability to detect and help prevent a failure mode before it impacts the business.

A critical success factor to assist with developing effective detection controls will be 'co-creation' across the eco-system. Zero Business Impact 'Application Availability' is the ultimate goal, which can't be achieved without collaboration amongst partners, suppliers and customers. This simply doesn't happen enough in today's 'demarcation-centric' contracts and support models.

The practical application of this framework is broken down into two areas:

- **Empirical application** – Telstra has run this framework on historical Incident and Problem data and identified the most common failure modes that cause network downtime. The two prevalent failure modes are 'unknown single points of failure' and 'configuration errors'. In this report, we explain how to identify and mitigate this risk.

- **Visionary application** – Technology such as SD-WAN & NFV are expected to assist with improving quality and reducing risk, however, we don't know the foreseeable failure modes using this new technology due to a lack of observed data. Our framework encourages the anticipation of failure modes and development of agile detection controls across the life cycle.

Author

Richard Thomas, Head of Service Management, South Asia, richard.thomas@team.telstra.com

International Sales and Service, Telstra Enterprise

[1] Based on average annual availability from a sample of 20 Telstra customers in 2018. Sites with a resilient design only.

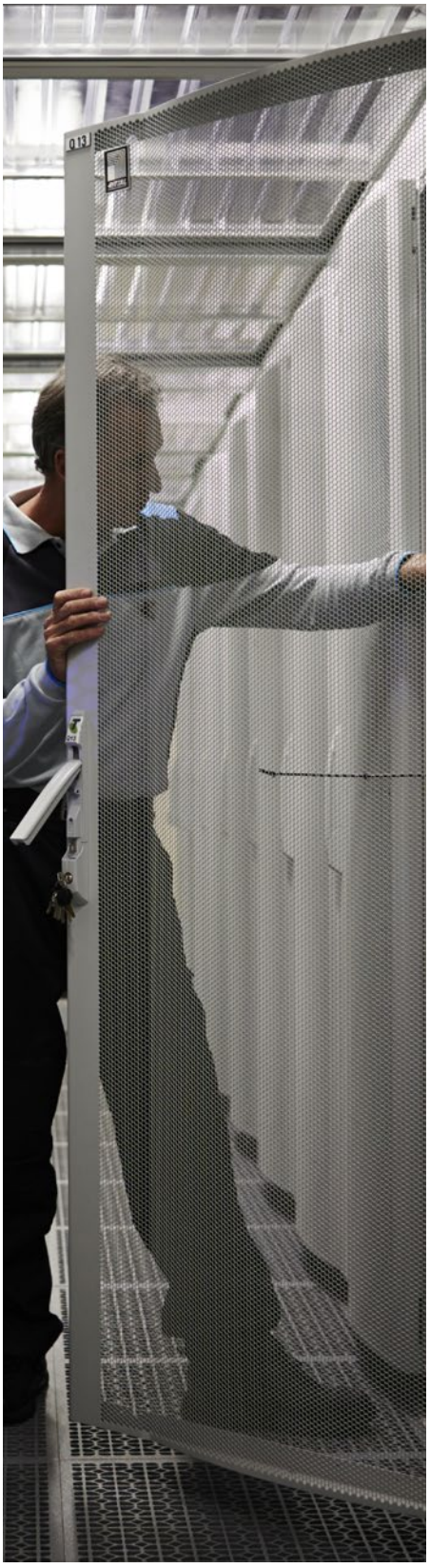


Table of Contents

Introduction 04

Empirical Contributors of Network Downtime 06

The Journey Towards Zero Preventable Downtime - Ideate to Release 10

The Journey Towards Zero Preventable Downtime - Operate 12

Ruthless Prioritisation 14

Potential Return on Investment 16

Conclusion 18

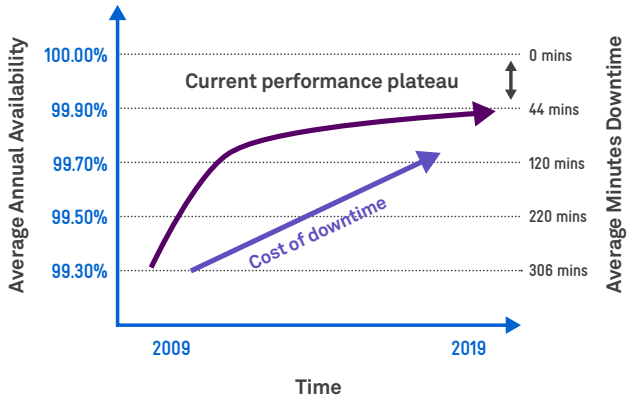
Introduction

Introduction

Organisations today have zero tolerance towards unplanned network downtime. In the era of digital transformation, network downtime has an unprecedented impact on enterprises with negative ramifications on revenue, productivity and brand image. The financial impact of one hour of Downtime is over USD 300,000 according to Gartner[2]. However, this number is an average and for large financial institutions and tech giants, we hypothesise these numbers can reach well over USD1M[3] per hour of Downtime.

With this in mind, have IT Service Management teams materially reduced network downtime over the past decade? Our hypothesis is that the improvement has been marginal and network availability during this period has plateaued to range from 99.00%-99.99%.

Figure 1: Business Impact of Downtime vs Network Availability Performance, 2019



There are many reasons for this plateau in availability levels. However, we believe the main contributing factors to be:

Ineffective prioritisation of risk

Lack of co-creation and transparency within the ecosystem leading to unknown Single Points of Failure

An imbalanced focus on firefighting current problems rather than anticipating future problems

The Journey towards Zero Preventable Downtime – Unwrapping the Title

Given the increased reliance on the Network in the Digital Economy, the Telstra Global Service Management team has developed the **'Journey Towards Zero Preventable Downtime'.** Each word in the title of this framework has been carefully crafted as explained below.

Journey Towards

The resources, ambition and determination to move to 100% Application Uptime presents a formidable challenge and one that requires a mindset change from all members of the ecosystem. It is a call of action to the entire ecosystem to put our customers first and attempt to significantly increase transparency while remaining cognisant of security and intellectual property concerns.

The journey also represents the mindset shift that many incidents and problems can potentially be prevented at the design, and in particular the transition phase.

Zero

Our ambition is aligned to the fact enterprises today have **Zero** tolerance for downtime.

Preventable Downtime

Preventable downtime has been defined as follows - *"upon completing root cause analysis on any Incident creating Downtime, could it have been prevented not withstanding any extreme commercial, environmental and operational constraints?"*

This standard will be subject to continual service improvement as what is considered 'unpreventable' should be periodically challenged and ultimately the framework aims to be titled "Zero Downtime".

[2] The Cost of Downtime, Gartner Blog Network, 2014
[3] For obvious reasons, large organisations do not publish their cost of downtime, however, incidents such as the following clearly have the potential to cost in excess 1M in lost revenue. Reference: Visa outage hits payments across Europe, Finextra, 2018

Empirical Contributors of Network Downtime

Empirical Contributors of Network Downtime[4]

While there are a number of factors that contribute towards IT or application downtime, often the highest contributor is the **Network.** Deeper analysis shows that 80%[4] of Network related incidents[4] that actually cause Downtime or Site Isolations are caused by:

- Unknown Single Points of Failure (SPOF)
- Power outages
- Configuration errors
- Defective hardware

Figure 2a: Common contributors of IT downtime

1. Network outages
2. Human error
3. Server/storage failures
4. IT power outages
5. Capacity constraints (usage spikes/surges)
6. Natural disasters or weather events
7. Third-party supplier or cloud outages

Figure 2b: Primary contributors of sites isolated due to a network outage[4]

The underlying causes for each of the top four contributors are presented below for customer sites with a resilient design, thus have dual network connections and customer premise equipment (CPE).

Figure 3: Primary root causes for network downtime (resilient site topology)

Unknown SPOF **40%**	Configuration Errors **20%**	Power Outages **20%**	Defective Hardware **10%**
Last-mile SPOF	Failover configuration failure	No UPS	Unknown LAN SPOF
Recovery time	WAN routing fault	UPS failure	Unknown WLAN SPOF
Subsea cable SPOF	(W)LAN configuration fault	Power surges	

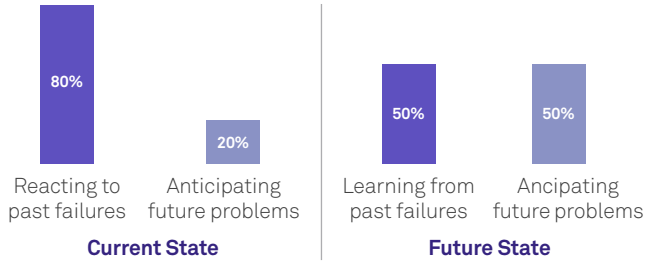
Our risk management framework practices the principle of 'ruthless prioritisation'. The top two failure modes, which cause over 50%[4] of Network downtime – **Unknown Single Points of Failure** and **Preventable Configuration Errors** – would help provide the greatest possible initial return on investment. Once the risk ranking per site for these two failure modes has materially reduced, organisations should shift their focus to the next highest ranked failure mode.

[4] Root causes from a sample of 20 Telstra customers in 2018. Data summarized and rounded to nearest 10%. Data only included incidents that (a) involved a resilient site isolated and (b) involved complete isolation from the network.

The Risk Management Framework on The Journey Towards Zero Preventable Downtime

Today, IT Service Management organisations are still allocating disproportionate levels of resources toward 'fire fighting' [5] and problem management procedures, which despite delivering significant value are inherently reactive approaches. The Journey towards Zero Preventable Downtime seeks to drive a proactive approach and was developed around the central theme of **'anticipating future problems to solve today'.** In this regard, the framework was essentially created as a network downtime risk identification, prioritisation and mitigation model.

Figure 4: Illustrative view of today's focus of IT Service



Our risk management tool uses a simplified adaptation of **Failure Mode Effects Analysis (FMEA),** which was created in the 1940s by the US military. FMEA is a systematic approach for identifying and mitigating possible failures for a product or service. Historically, FMEA has been used as a tool to help achieve 'zero defects' in aerospace and automotive industries.

Traditionally IT & Network centric organizations have taken a Project-Orientated Management approach, often with distinct design, deliver/transition and operate phases. There is momentum building to replace this approach with Product-Orientated Management focus. There is one very important principle of a Product-Orientated Management focus that has been applied to the framework relating to the treatment of Risk. Mik Kersten, in the book 'Project to Product' (2018) describes the treatment of Risk between the approaches as follows:

Project-Orientated Management	Product-Orientated Management
Delivery risks, such as product/market fit, is maximized by forcing all learning, specification, and strategic decision making to occur up front	Risks is spread across the time frame and iterations of the project

In the past, the majority of the effort to identify risks related to Network Downtime was at the front-end of the project for example the design stage. Identifying, prioritising, and helping to mitigate risk periodically from the Ideate/Create (or design) to Release (or deliver/transition) and finally the operate stage is a key principle of The Journey Towards Zero Preventable Downtime Framework.

[5] This position is an educated observation and not substantiated by empirical data.

The Journey Towards Zero Preventable Downtime Risk Management Framework (Continued)

Figure 5: The Risk Management Framework

Identify Failure Modes	Creative identification of all possible failure modes
Identify Failure Effects	Assess the effects of identified failure modes and their potential impact to downtime
Potential Causes	Identify the potential causes of the failure modes
Detection Controls*	Create detection controls to mitigate incidents **before** occurrence
Rate Severity	Rate the severity of each failure mode from **1** (low) to **5** (catastrophic)
Rate Occurrence	Rate the occurrence of each failure mode from **1** (low) to **5** (high)
Rate Detection	Rate the detection controls for each failure mode from **1** (certain) to **5** (none)
Risk Prioritisation	Severity X Occurrence X Detection = Risk Ranking Score (used for prioritisation)
Identify Mitigation Plans	Identify targeted mitigation plans to help reduce risk
Ongoing Monitoring	Risk ranking scores enables demonstrable quantification of potential risk reduction

***Detection Controls** differentiate this approach from best practices such as ITIL. In this adaptation of FMEA, we use Detection as the rating number corresponding to the likelihood that the current detection-type controls will assist in detecting the failure mode/cause before it adversely impacts the business.

Table 1: Sample Legend for Preventable Downtime Risk Ratings

SEVERITY RATING		OCCURRENCE RATING		DETECTION RATING	
Rating	**Description**	**Rating**	**Description**	**Rating**	**Description**
5	Country/DC Network isolation	5	Failure mode occurs **weekly**	5	**No** detection controls in place
4	Business Critical Site isolated	4	Failure mode occurs **monthly**	4	**Low** probability of detection before any incident
3	Large Site >50 users isolated	3	Failure mode occurs **quarterly**	3	**Medium** probability of detection before any incident
2	Small Site <50 users isolated	2	Failure mode occurs **annually**	2	**High** probability of detection before any incident
1	Remote Site <5 users isolated	1	Failure mode occurs **rarely**	1	**Certain** detection before any incident

The Journey Towards Zero Preventable Downtime - Ideate to Release

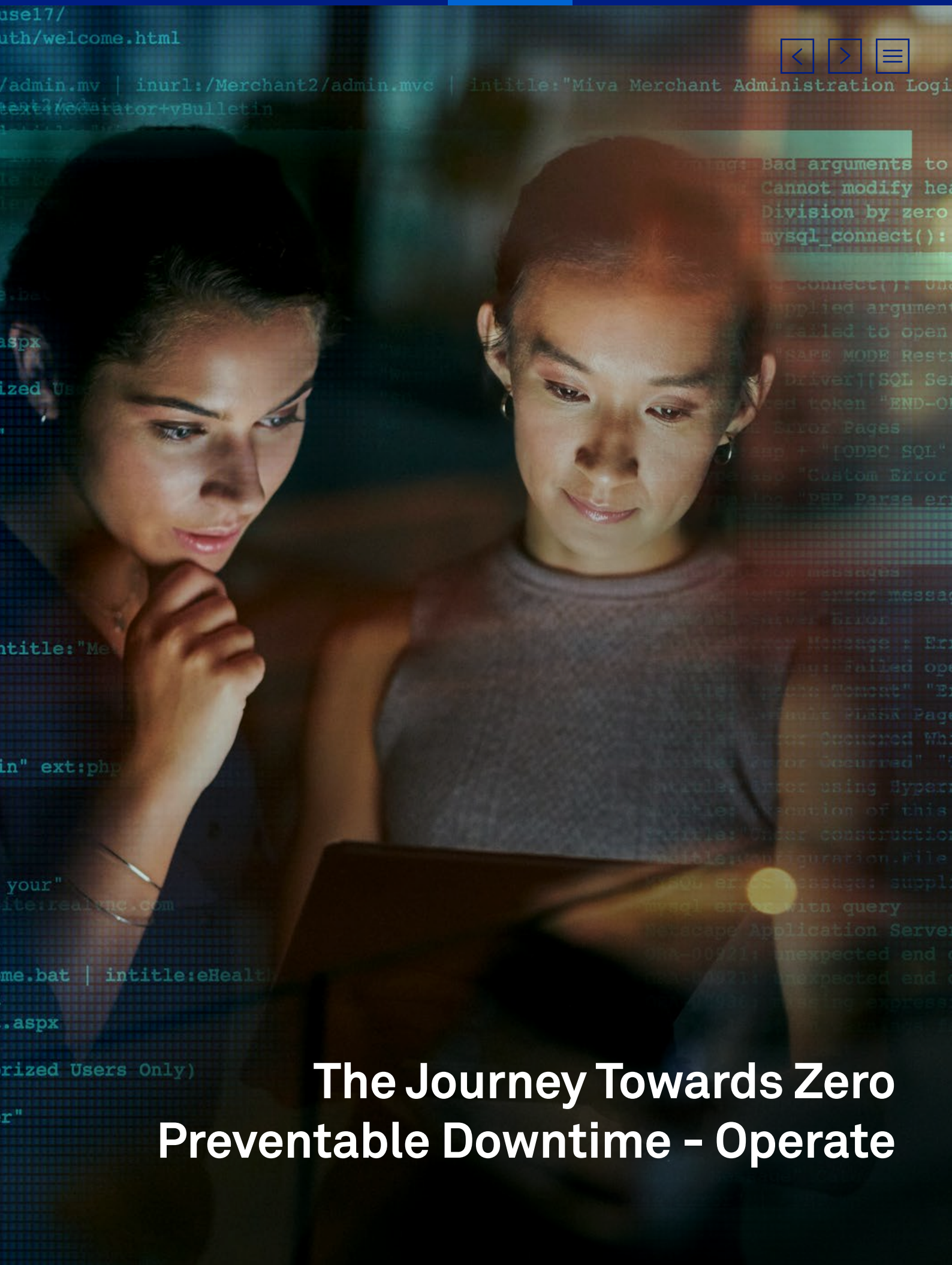
Ideate Create Release Operate

The Journey Towards Zero Preventable Downtime - Ideate to Release

From the flow of Ideate to Release the two most common failure modes[6] for Network Downtime (i.e. Unknown Single Points of Failure and Preventable Configuration Errors) are used as an example below.

	UNKNOWN SINGLE POINT OF FAILURE	PREVENTABLE CONFIGURATION ERRORS
IDENTIFY FAILURE MODES	Both last mile Access Circuits terminate into the same local exchange.	Core router class of service configuration misaligned to edge router.
IDENTIFY FAILURE EFFECTS	Failure at SPOF location would isolate the site from the network.	During congestion, voice, video and business application performance will be degraded.
POTENTIAL CAUSES	(1) Local diversity not ordered by customer or supplier. (2) Lack of transparency on path.	Incorrect configuration applied to core PE routers and or edge CE router during commissioning.
DETECTION CONTROLS	(1) Include a last mile path transparency clause in contracts. (2) Create delivery task to validate and obtain evidence.	Design team create a robust 'Validation and Testing Plan' to hand over to the delivery team.
RATE SEVERITY	High severity as site would be isolated. Ranking "4" or "5" based on criticality of site.	High severity as key applications could suffer. Ranking "4" or "5" based on criticality of site.
RATE OCCURRENCE	Occurrence relative to geography[6] e.g. Singapore ranked at "1" versus remote sites in India likely ranked at "3".	Typically, a low occurrence failure mode ranked at "1".
RATE DETECTION	Detection rating typically very poor, e.g. ranked at "5" due to a lack of transparency in the industry.	Detection at Design and Delivery stage generally high, e.g. ranked at "1" or "2".
RISK RANKING	5 (Severity) X 3 (Occurrence) X 5 Detection = 75. High ranking risk.	5 (Severity) X 1 (Occurrence) X 2 Detection = 10. Low ranking risk, but high Severity.
IDENTIFY MITIGATION PLANS	Design team to provide clear SoW to Delivery team to execute recommended detection controls.	Design team to provide clear SoW to Delivery team to execute recommended detection controls.
ONGOING MONITORING	Risk transferred over to 'Operate' to monitor.	Risk transferred over to 'Operate' to monitor.

[6] Root causes from a sample of 20 Telstra customers in 2018. Data summarized and rounded to nearest 10%. Data only included incidents that (a) involved a resilient site isolated and (b) involved complete isolation from the network.



The Journey Towards Zero Preventable Downtime - Operate

The Journey Towards Zero Preventable Downtime - Operate

At the heart of the Journey Towards Zero Preventable Downtime is the third phase of the risk management framework, which is operationalised through a periodic audit program run by the Telstra Global Service Management team. As stated earlier, the initial focus is to tackle two failure modes that typically cause over 50% of network downtime or site isolations at sites designed to be resilient – namely, Single Points of Failure and configuration audits. These elements have been incorporated into the audit framework as described below.

Periodic Audit Framework for the Journey Towards Zero Preventable Downtime

The periodic audit framework was designed in accordance with the risk management framework to help achieve the following:

1. Facilitate 'ruthlessly effective' **risk prioritisation**
2. Re-allocate more focus on **anticipating** future problems
3. Significantly enhance the **transparency** our customers have on their network risk

Figure 6: The salient elements of the Periodic Audit Framework

Inventory
A full inventory listing of customer sites including details such as circuit ID, SLA, access speed, port speed and CoS allocations.

Physical
Intense end-to-end analysis to ensure diversity of resilient paths (customer building – last mile – core – subsea cable – core – last mile – customer building).

Logical
Focused sub-set of preventive configuration audits on errors that can impact availability or performance (e.g. latency, jitter, packet loss, CoS, throughput).

Capacity
Review of capacity to ensure Applications meet end user expectations during normal operating conditions and during any failover to redundant connections.

Facility
Review of building at site to ensure there is no unknown SPOF (ex: dual risers, dual power, and UPS backup, fire-fighting system).

Risk Tracking and Mitigation
All failure modes are documented, rated, and mitigating actions formulated from design, delivery and periodically during the in-life contract period.

Ruthless Prioritisation

Ruthless Prioritisation

Service Management resources are finite and constantly under pressure to reduce their overall 'Cost to Serve'. In this regard, the ruthless prioritisation of service management resources is critical to help generate the greatest possible return on investment. The illustration below outlines how our framework can help to solve this problem.

Baseline Top 10 Failure Modes[7]

TYPICAL FAILURE MODES IN A MEDIUM SIZE NETWORK	Risk Rating Sev	Occ	Dec	RR
Last Mile SPOF	5	2	4	40
WAN Failover Configuration Error	5	2	3	30
LAN Failover Configuration Error	5	1	3	15
Recovery Time of resilient circuit	5	1	3	15
Power Surges	5	1	3	15
Subsea Cable SPOF	5	1	2	10
UPS Failure	5	2	1	10
Unknown LAN SPOF	5	1	2	10
CoS / BW Configuration Error	3	1	3	9
Unknown WLAN SPOF	3	1	2	6

Baseline Position

- Focus on highest-rated failures modes
- Given resource constraints, strategy for next six months could be:
  - Reduce last mile SPOFs by increasing end-to-end transparency
  - Audit and correct LAN and WAN fail-over configuration

Network Availability: 99.6%

Bz Impacting Downtime (hours): 3 hrs

Cost of Downtime (USD/Hr): 0.8M

Baseline + 6 months - Top 10 Failure Modes[7]

TYPICAL FAILURE MODES IN A MEDIUM SIZE NETWORK	Risk Rating Sev	Occ	Dec	RR
SD-WAN orchestration layer error	5	1	5	25
Recovery Time of resilient circuit	5	1	3	15
Power Surges	5	1	3	15
Subsea Cable SPOF	5	1	2	10
UPS Failure	5	2	1	10
Unknown LAN SPOF	5	1	2	10
↓ Last Mile SPOF	5	1	2	10
↓ WAN Failover Configuration Error	5	1	2	10
↓ LAN Failover Configuration Error	5	1	3	15
CoS / BW Configuration Error	3	1	3	9
Unknown WLAN SPOF	3	1	2	6

Baseline Position + 6 months

- Plans to reduce Last Mile SPOFs and WAN/LAN configuration errors were very successful e.g. several last mile SPOFs identified and mitigated by deploying a 4G back-up. Risk rating updated accordingly
- Plans to roll out SD-WAN in the next six months have unsurfaced a new failure mode with the highest ranking
- Resources can focus again on the top three failure modes to derive the greatest ROI

Network Availability: 99.9% ↑

Business Impacting Downtime (hours): 1 hr ↓

Cost of Downtime (USD/Hr): 1M ↑

[7] This example uses data that is purely illustrative, however, was inspired by real customer experience.

Potential Return on Investment

Potential Return on Investment

One of the most important and underutilised value-based measures for IT Service Management is the following:

(Downtime Cost Prevented) - Cost to Serve) ÷ Cost to Serve = ROI

- To derive 'Downtime Cost Prevented', multiply the number of 'downtime hours' prevented by the average cost per hour of downtime.
- To calculate a monthly 'Cost to Serve', add direct costs associated with the ITSM wrap e.g. loaded resource cost, systems and tools, etc.
- We recommend to cap the 'downtime hours' at seven hours per critical site per month (approx. 99.0% availability. This allows the exclusion of the infrastructure costs necessary to meet the target service level % e.g. network, servers, etc). At Telstra, we are currently modelling capping at 99.9% (our most common service level). When we have tangible evidence, agreed by our customer, that a material risk(s) was mitigated, we may consider crediting the Service Manager with 45 minutes of 'Downtime Prevented' against the occurrence modelled for this related failure mode e.g. if Occurrence was modelled at one in 12 months, one month credit would be allocated. This would then be multiplied by the agreed 'cost of downtime' relevant for that particular customer.

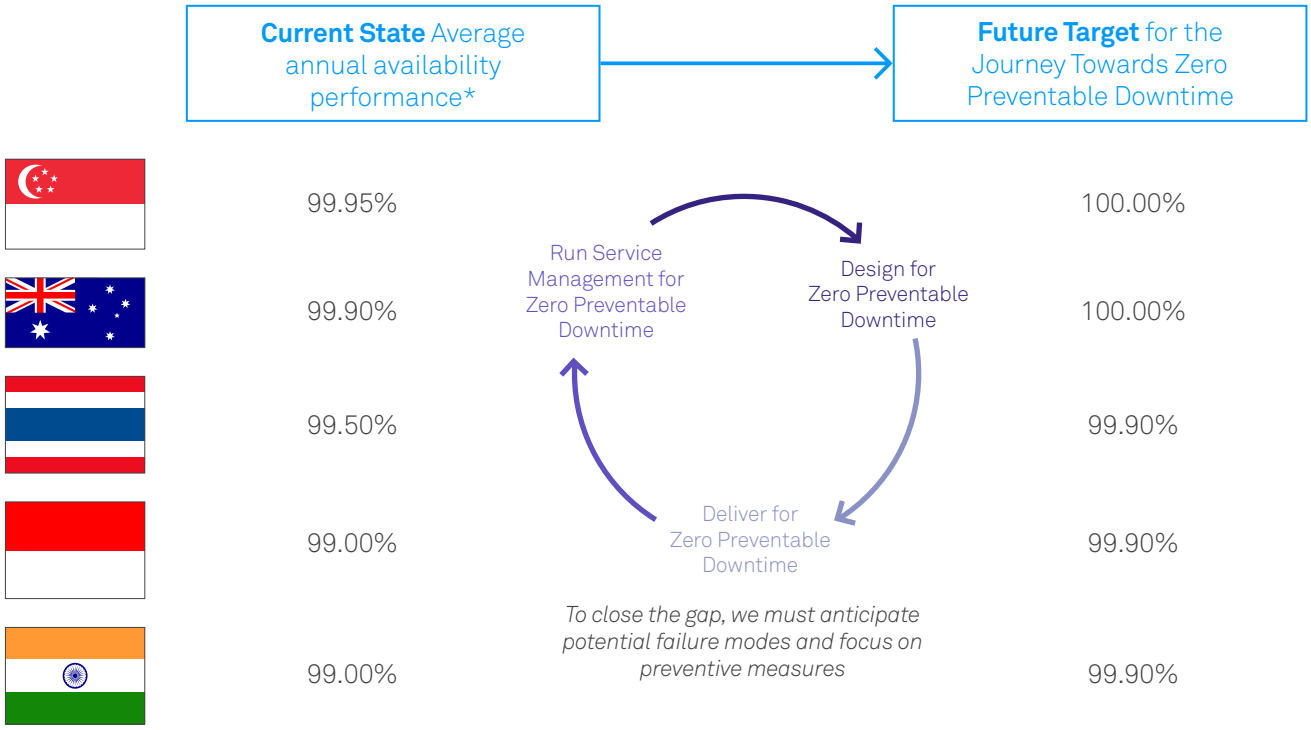
Conclusion

Conclusion

Today's focus is crystal clear. The majority of preventable network downtime is coming from two sources (1) Unknown Single Points of Failure and (2) Preventable Configuration Errors. In this connection, the immediate focus on our audit framework is to deliver positive incremental results, by enhancing organisational visibility of network vulnerabilities.

Conceptually this framework can be summarised using the illustrative model below.

Figure 7: Current State vs. Future Target of the Journey Towards Zero Preventable Downtime



*Sample size of 100 sites per country with a resilient site topology. Mixture of sites in metropolitan areas (80%) and remote areas (20%).

As digital transformation and digital disruption gather momentum at an exponential rate, the failures modes of the future are unknown. The key factor that sets this framework apart from traditional approaches used today is its forward-looking and proactive nature.

If you are interested to learn more about the application of this approach, please contact the following:

International

Asia +852 2983 3388 **Americas** +1 877 835 7872 **EMEA** +44 20 7965 0000 **Australia** +61 2 8202 5134

Richard Thomas richard.thomas@team.telstra.com