Skolkovo Institute of Science and Technology

A FRAMEWORK FOR ARCHITECTING FEDERATIONS OF SYSTEMS

Doctoral Thesis

by

IGNASI LLUCH I CRUZ

DOCTORAL PROGRAM IN ENGINEERING SYSTEMS

Supervisor

Professor Alessandro Golkar

Moscow - 2017

This page is intentionally left blank

Abstract

This thesis presents a quantitative framework to assess the conditions for the formation and evolution of Federations of Systems (FoS). Federations are a type of Systems of Systems (SoS) where the component systems, or peers, present independent goals, management and operations. The systems in a federation cooperate for mutual benefit, often tapping into underutilized resources. These features are not unique to engineering systems, being also recognizable in peer-to-peer access-based markets, or the *sharing economy*. Hence this thesis analyzes both engineering and non-engineering systems.

Nowadays, the potential for cooperation between systems lies on operations on digitalized data, such as context information sharing, distributed processing, data relay and bandwidth sharing. In the field of economic activities, cooperation can encompass a broader set of capabilities in transportation, accommodation, or finance. Federations are becoming a reality due the widespread adoption of network communication technologies and the computing power available in today's systems. Unfortunately, the literature lacks dedicated analysis of FoS. This thesis bridges this gap for the first time.

Due to the voluntary nature of the cooperation in an FoS and the independency of the parties involved, system architects designing a system in such an environment face additional difficulties. The architectural decisions to be taken are coupled with other system's architectures; its benefit depends on uncertain future scenarios, and the value delivered by the system through its life is exposed to the dynamics of the FoS. Since architecting a system to federate may increase the costs due to the additional interfaces needed, the cost-benefit trade-offs must be carefully characterized in order to support the architect's decisions.

This thesis develops a cohesive framework addressing these issues, based upon the concept of synergy, defined as the aggregated benefit of cooperation between systems. Using synergy as a stepping stone, this work builds a framework based upon the tradespace exploration paradigm and Markov Decision Processes (MDP) to predict the emergence and evolution of federations of engineering systems. Using the framework, this work analyzes Federated Satellite Systems (FSS), ridesourcing services and Wireless Community Networks (WCN).

In the case of FSS, we demonstrate that a federation of Low Earth Orbit (LEO) Earth Observation (EO) satellites can be beneficial for the participants, improving mission data latency and enhancing the cost-effectiveness of the missions' downlink bandwidth. In particular, we show that federating is advantageous when the FSS lifecycle interface costs are below 15 MUSD. FSS interfaces consist of an Inter-Satellite Link (ISL) and data handling equipment. Moreover, we identify scenarios

where cooperating is already beneficial from a 2-satellite federation. We analyze the sensitivity of federation benefits to a wide array of parameters, including interface costs, discount rates, architectural cost constraints, and constraints on cooperation. We conclude that the advantages of federating can amount up to 40% of a mission's ground segment and communications subsystems budget.

This thesis then analyzes the *ridesourcing* market of New York (NY) between 2012 and 2017, interpreting it as a federation of drivers and riders. This case serves as a retrospective validation case. Using real data of the NY taxi market, the framework satisfactorily predicts and matches the evolution of fares, fare rates, driver earnings, number or drivers, wait times and earnings of the *UberX* service.

Finally, the study in WCNs shows their potential to bridge the global rural digital divide. We show each peer can save from 10 to 60 USD monthly for their internet access respect to other solutions like satellite-provided internet and fixed infrastructure.

Besides the practical insights mentioned, this work also derives general observations. Based upon the case study results, we examine the nature of the advantages of federation. We identify they originate in either functional emergence or a cost-effective resource reallocation. Furthermore, we identify and illustrate 3 emergence modes and 2 failure modes of federations. We finish with recommendations on FoS governance aimed to favour emergence and avoid FoS collapse. This thesis contributes at a methodological and practical level to the study of SoS and FoS.

Acknowledgements

That's for sure the section that I've dreamt to write for years now. Well, here we are. If somebody would have told me, only 5 years ago, you will have a PhD from a university in Moscow, I might have believed it, but nevertheless would have thought *'now that's an interesting turn of events'*. We can attribute this happening largely to my advisor and mentor, Alessandro Golkar. Not only he convinced me to join a brand new, freshly created university, but then he proceeded, on the following 4 years, to build a space center from scratch and develop our group to an internationally competitive research team. Now add to this the inherent challenges of operating within a newly established institution that changes campus, leadership and facilities as it grows, and you can start to see the extent of this achievement. I'm proud and grateful of the role that I could play in this while working with him. Alessandro is a person who makes things happen; from all lessons I've learned from him, the ambition and strive for excellence are probably the most important.

I am also very thankful to my colleagues these years. Hripsime and Dominik joined the team 3 years ago and I do not see myself in Skoltech without them. They've provided key suggestions and help both in the academic and personal life. Being an expat in Moscow has its moments and they've saved me from sleeping under a bridge a couple of times. I'd also like to thank Dima, a true entrepreneurial spirit. I am proud I've seen Rustam grow from a student to a professional and to a PhD candidate, and helping him in his thesis and flight campaigns is one of my most happy memories in Skoltech. I would also like to thank Olga, Udri, Simone, Raffaele, Christianna, Dominique, Gianni, and Armen for bearing with me.

Also many thanks to the reviewing team, first to Nina for helping with the thesis and everything else. Thanks again Hripsime, Dominque, Dima, Dom and Raffaele for going through it.

Some of the Skoltech staff can make your day a little bit brighter. I'd like to thank Natalia, Yana, Ksenia, Ido and Anastasia S. for making my life easier this years, and having the energy to be charming at all times. I am also grateful to the professors and academics that helped making this thesis better, and encouraged me through the PhD, including Clement Fortin, Rupert Gerzer, Zeljko Tekic and Kelvin Willoughby. Last but not least, I am always impressed and grateful at how my parents are supportive through all of my endeavors. I can say, 'I'm leaving for work in the Netherlands', or U.S., or for a PhD in Russia, or 'I'm going to start a company', and they are up for it instantly.

Happy reading!

Contents

List of Figures

List of Tables

List of Acronyms

ADSL	Asymmetric Digital Subscriber Line
AP	Argument of Perigee
AP	Access Point
BATMAN	Better Approach to Mobile Ad-Hoc Networking
BGP	Border Gateway Protocol
BOL	Beginning of Life
CAPEX	Capital Expenses
CAR	Cumulative Abnormal Returns
CAS	Complex Adaptive Systems
CN	Community Networks
COTS	Commercial-of-the-Shelf
DoD	Department of Defense
DoE	Design of Experiments
DSS	Distributed Space Systems
DTMC	Discrete-Time Markov Chain
EO	Earth Observation
EOL	End Of Life
FCC	Federal Communications Commission
FFCS	Free-Float Car Sharing
FO	Federation Options
FoS	Federation of Systems
FSS	Federated Satellite Systems
GB	Global Benefit
GC	Global Cost
GEOSS	Global Earth Observation System of Systems
GN	Gastner-Newman
GPS	Global Positioning System
GR	Geostationary Relay
GS	Ground Station
GSA	Ground Segment Architecture
HOV	High Occupancy Vehicle

HSN	Hetereogeneous Spacecraft Networks
ICT	Information and Communications Technologies
IEEE	Institute of Electrical and Electronics Engineers
INCOSE	International Council on Systems Engineering
IoT	Internet of Things
ISL	Inter-Satellite Link
ISP	Internet Services Provider
ITU	International Telecommunications Union
JAXA	Japan Aerospace Exploration Agency
LB	Local Benefit
LC	Local Cost
LCS	Littoral Combat Ship
LEO	Low Earth Orbit
LP	Linear programming
LTE	Long-Term Evolution (a communications standard)
M2M	Machine-to-Machine
MA	Multiple Access (a TDRSS mode)
MANET	Mobile-Ad-Hoc Networks
MATE	Multi-Attribute Tradespace Exploration
MAUT	Multi-Attribute Utility Theory
MDO	Multi-Disciplinary Optimization
MDP	Markov Decision Process
MOO	Multi-Objective Optimization
MS	Minimum Sharing (constraint)
MUSD	Millions of United States Dollars
NASA	National Aeronautics and Space Administration
NBC	Net Benefit of Cooperation
NBI	Normal Boundary Intersection
NEN	Near Earth Network
NYC	New York City
OBDH	On-Board Data Handling
OD	Origin-Destination
OPEX	Operational Expenses
OSLR	Optimized Link-State Routing

PNB	Perceived Net Benefit
RAAN	Right Ascension of the Ascending Node
RoI	Return on Investment
RSC	Responsive Systems Comparison
SA	Single Access (a TDRSS mode)
SAI	Systems Architecting with Ilities
SAR	Synthetic Aperture Radar
SB-WSN	Space-Based Wireless Sensor Networks
SDR	Software-Defined Radio
SETI	Search for Extra-Terrestrial Intelligence
SGR	Space-to-Ground Rate
SMA	Semi-Major Axis
SoS	System of Systems
SoSTEM	SoS Tadespace Exploration Method
SPF	Stand-alone Pareto Front
SQP	Sequential Quadratic Programming
SSO	Sun-Synchronous Orbit
TDN	Time-expanded Decision Networks
TDRSS	Tracking and Data Relay Satellite System
TLC	Taxi and Limousine Commission
TNC	Transportation Network Companies
V2G	Vehicle to Grid
V2V	Vehicle-to-Vehicle
VWFO	Value-Weighted Filtered Outdegree
WCN	Wireless Community Network
WiMAX	Worldwide Interoperability for Microwave Access

Chapter 1 Definitions

Despite the definitions here, the concepts introduced below are subject to closer scrutiny and explanation through this document, especially in the literature review and approach chapters.

Cooperate: To associate with another or others for mutual benefit (Merriam-Webster).

Collaborate: To work jointly with others or together especially in an intellectual endeavour (Merriam-Webster).

Emergence: “What appears, materializes, or surfaces when a system operates. […] Most obviously and crucially, function emerges” (Crawley et al., 2016a).

Engineering system: A class of systems characterized by a high degree of technical complexity, social intricacy, and elaborate processes, aimed at fulfilling important functions in society (De Weck et al., 2012).

Function: “A function is what a system does; it is the activities, operations and transformations that cause, create, or contribute to performance. […] Function is the action for which a thing exists or is employed” (Crawley et al., 2016a).

Form: “Is what a system is, the physical or informational embodiment that exists. Form has shape, configuration, arrangement or layout. Form […] is necessary to deliver function” (Crawley et al., 2016a).

Stakeholder: Individual or organization having a right, share, claim, or interest in a system or in its possession of characteristics that meet their needs and expectations (“ISO/IEC/IEEE International Standard – Systems and software engineering – System life cycle processes,” 2015).

Synergy: A mutually advantageous conjunction or compatibility of distinct business participants or elements (as resources or efforts) (Merriam-Webster).

System: 1) A set of entities and their relationships, whose functionality is greater than the sum of functional entities (Crawley et al., 2016a).

2) A construct or collection of different elements that together produce results not obtainable by the elements alone (NASA, 2007).

Chapter 2 Introduction

The digital revolution and the advent of the information age (Castells, 2011) has deeply transformed the economic processes and social interactions of the world we live in. Over the last 50 years, we have witnessed ground-breaking advances in Information and Communications Technologies (ICT), the miniaturization and embedding of computing power, and the ascent of the Internet. On the last decade alone, we have seen the widespread adoption of smartphone devices, mobile connectivity and the availability of virtual computing resources on demand (Aldhaban, 2012). This has deeply changed many social interactions, organizational procedures and entire business markets.

Hence, most of today's engineering systems include, or *are*, ICT. The potential of the ICT *within* engineering systems has been long realized and will continue to be implemented. Instead this work explores the application of these technologies as means to support cooperation *between* independent, operational systems. That is, how to enhance the value delivery of engineering systems through cooperation, in a time of economic distress and political uncertainty (Stiglitz, 2009) where making engineering systems more cost-effective is fundamental. To address this general goal, this research adopts a systems' architecting perspective (Crawley et al., 2016a). This thesis identifies issues related to systems' cooperation in architecting terms, and proposes quantitative approaches to evaluate these issues. We start by examining what is the potential for cooperation between engineering systems.

2.1 Cooperation between engineering systems

We explore here systems such as transportation vehicles, sensory networks (Karl and Willig, 2007), industrial production systems or spacecraft to identify the advantages of cooperation and the means to implement it.

The means, or technical infrastructure for system-to-system cooperation, would seem to be falling in place. Paradigms like the Internet-of-things (IoT) (Fleisch, 2010) and Machine-to-machine (M2M) (Holler et al., 2014) have emerged due the miniaturization advances in sensing and processing, and the appearance of methods to control and harness the full potential of interconnected systems –machine learning, big data analytics–. These concepts stimulate interface standardization and interoperability efforts which are a prerequisite to the cooperation we envision here.

Let us now address the potential advantages. What types of cooperation opportunities can systems embrace using IoT/M2M technologies? With the current technological forms, these opportunities are embodied by operations on digitalized data: data relay from origin to end destination (Yu et al., 2011), storage for others (Armbrust et al., 2010), distributed processing (Gunter and Maessen, 2013), and context-information sharing for enhanced operational awareness (Festag, 2014). These applications are indeed very common operations on the existing computer network infrastructure, and usually performed under the cloud computing model (Armbrust et al., 2010). Engineering systems for which such operations have been theorized include road vehicles and notably federated satellite systems (FSS) (Golkar and Lluch i Cruz, 2015).

FSS is based upon a space communications network (Lluch et al., 2015), used by participant satellites to exchange resources, such as bandwidth and computing power, for mutual benefit. FSS participants cooperate on an opportunistic, ad-hoc basis. Satellites in the federation trade resources on a commercial basis, to increase the cost-effectiveness of their capabilities and/or improve their performance. Figure 1 illustrates this concept.

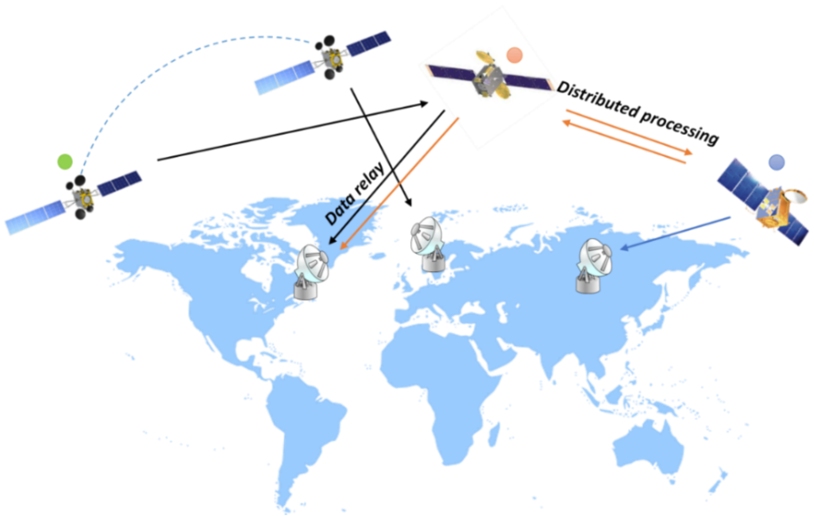

Figure 1. The concept of Federated satellite systems, where missions support others to process and downlink data.

This thesis generalizes the term ‘federation’, applying it to engineering systems like road vehicles and community networks (Vega et al., 2012). Federations of systems (FoS) are then type of System of Systems (SoS), more precisely a virtual SoS (Maier, 1998), as described in detail in Chapter 4. In this work, we say two or more systems are federated if they cooperate under a common agreement, while retaining their managerial, operational and goals independence. By extension, a federation is a set of systems voluntarily cooperating for mutual benefit. These definitions are discussed in detail in Chapter 5.

In the light of our discussion here, the cooperation between systems is understood as a resource exchange to improve the performance and/or cost-effectiveness of the participant’s operation. This draws parallels with a recent trend in economic activities, the *sharing economy*. We examine it next to identify and draw any key lessons for engineering systems.

2.2 Lessons from the sharing economy

The last 5-10 years have seen the rise of business models based on accessibility and peer-to-peer exchange, or informally, *the sharing economy* (Dervojeda et al., 2013). While the keyword here is *sharing* instead of *cooperation,* several elements of the sharing economy can be exported for our discussion in the engineering systems field.

The rise of the sharing economy is a result of the wide adoption of (mobile) internet connectivity and motivations of economic and social nature, namely the rising cost of living, the global economic crisis, decreased consumer trust in big corporations, and concerns about the environment (Dervojeda et al., 2013; van den Steenhoven et al., 2016). The sharing economy includes companies which match drivers with riders for peer-to-peer transport services (Chan and Shaheen, 2012), which list private dwellings for short-term accommodations (Zervas et al., 2016), crowdsource loans (Mollick, 2014) and offer workforce from individuals to individuals (Sundararajan, 2016). The current sectors are then transportation, retail, accommodation, service & labour, and finance (Johal and Zon, 2015).

The sharing economy is still an ill-defined term, often connected with other ambiguous terms like *collaborative economy,* the *gig economy,* and *peer economy* (Botsman, 2013). The fine details of the business models can vary from company to company. However, the basic conception is that the sharing economy encompasses companies which support peer-to-peer exchanges of underutilized assets or skills. This is the sense in which we refer to the sharing economy in this thesis. In the *sharing economy*, ICT supports a modern version of the old paradigms of communal sharing and bartering. The sharing economy, as depicted in Figure 2, is about matching a peer with an asset or skill, through a coordination platform, with a peer which has specific need on a specific timeframe. The latter compensates the supplier, and uses this system as an alternative to existing market solutions and exclusive asset ownership.

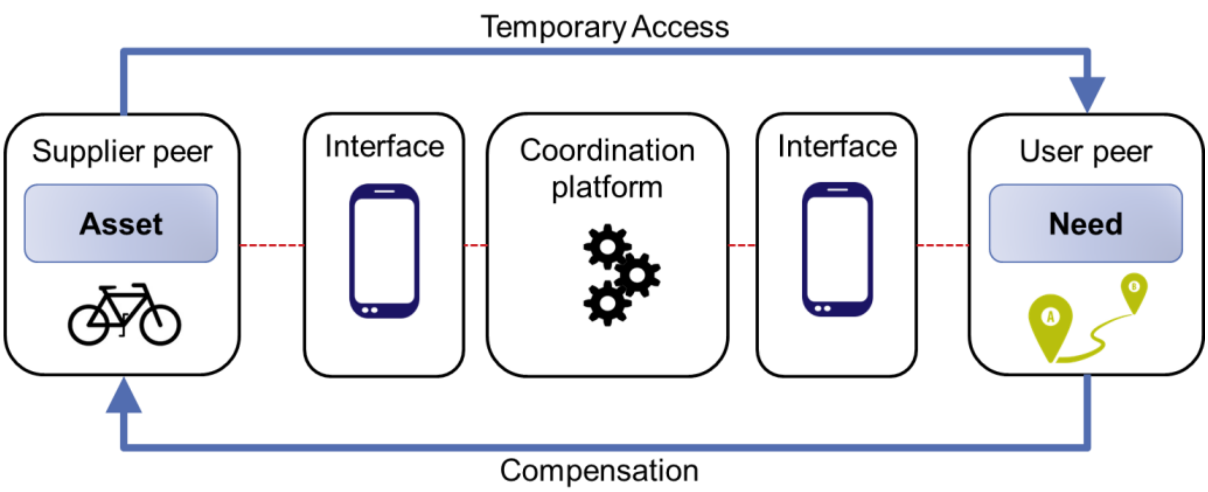

Figure 2. Basic elements and processes of accessibility-based peer-to-peer markets.

In this way, the sharing economy exploits underutilized assets such as cars, residences or savings and creates side-revenue of existing skills or capabilities, effectively distributing the cost of ownership. Can we apply these features to engineering systems?

While there are parallels, there are indeed some differences when applying these concepts to engineering systems or to economic activities. Table 1 clarifies the common ground and the differences between both. The dimensions considered for our comparative analysis are the assets or capabilities for lease, the nature of the peers, the interface used by the peers to access the coordination platform, the nature of the latter, the backbone communications infrastructure, the channels used to make the transactions effective, and the adoption problems with these paradigms.

Table 1. Cooperating for mutual benefit: a comparative between economic activities and engineering systems.

	Sharing economy markets			Engineering systems	
Examples	***Ridesharing* (*Lyft, Blablacar, Uber*)**	***Labor* (*taskrabbit, Sorted*)**	***Accommodation* (*airBnB, flipkey*)**	***Satellites (FSS)***	***Wireless Community Networks (WCN)***
Assets/ Capabilities offered	Transportation	Skills	House or room	Bandwidth, processing power	Bandwidth
Nature of peers	Riders and Drivers	Taskers and customers	House hosts and guests	Satellite system owners	Household owners
Interface on peers	Device with internet connection	Device with internet connection	Device with internet connection	Inter-Satellite Link	Wireless router
Backbone infrastructure	Fixed-line and LTE	Fixed-line and LTE	Fixed-line and LTE	Space network	Meshed network
Coordination mechanisms	Centralized	Centralized	Centralized	Centralized or distributed	Centralized or distributed
Execution channel	Rendezvous	Rendezvous	Rendezvous	Space network	Wireless network
Adoption problems	Network externalities, trust issues	Network externalities, trust issues	Network externalities, trust issues	Network externalities, interface adoption	Network externalities, interface adoption

The differences that stem from the comparison in Table 1 are mostly at the execution of the exchange, the technical interfaces, and adoption problems. Coordination platform refers to the technology and related standards the peers use to access a coordination platform. The latter is the software used to assign assets and match peers. In the sharing economy, coordination platforms are usually the key technology developed by the businesses. The interfaces and backbone infrastructure are assumed to be in place and not deliberately developed for the sharing economy market: peers have devices with mobile or conventional internet connection, and use a limited set of operating systems the coordination software needs to be interoperable with. For engineering systems, especially in the case of satellite systems, the interfaces and backbone infrastructure are not always in place, and create an explicit cost-benefit trade-off of enabling the cooperation. In sharing economy markets, the asset or capability is accessed often face-to-face, via a rendezvous of the involved peers. For engineering systems, due the nature of the assets offered, the backbone infrastructure and interfaces are also the channel where the transactions are realized.

While in the sharing economy coordination platforms tend to be centralized, as means for the fostering companies to control the pricing and earn profit on transactions, in engineering systems this depends on the actual peer-to-peer network topology. Both in Wireless Community Networks (WCN) (Oliver et al., 2010) and in FSS, the coordination mechanisms can be de-centralized, that is, the peer matching happens in an ad-hoc fashion, depending on who peers are able to directly connect to others.

Last but not least, some adoption problems are common to both worlds and some are unique. Network externalities (Katz and Shapiro, 1985) are a critical problem for both sharing economy and federations of engineering systems. However, in some areas this is more acute than others. In ridesharing the numbers of available cars and drivers is fundamental to the value of the platform, and this is contingent to the riders using it, challenging initial adoption and success of such platforms. In the labour and accommodation market this problem is less acute; some value to the user is created from the first adopters.

Adoption is also challenged by trust issues between peers. Trust issues are commonplace for all digital trade, and the sharing economy is no exception. This also plays a role for engineering systems. However in the later there is another dimension to be taken into account: the cost-benefit analysis of deploying the cooperation interfaces. The details of these adoption problems and their connection to the architecture of the systems involved are the motivation of this work.

2.3 Problem description

In a scenario where systems cooperate and face issues of network externality and uncertain cost-benefit, the systems architecting and design efforts must include additional dimensions. The value of such cooperation, the exogenous design influences (adoption of interoperability standards), the uncertainty on the status of the third-party systems and potential evolution of those through lifetime are among the new topics system architects and designers need to include in their analysis. Systems engineering (Walden et al., 2015) and architecting (Maier and Rechtin, 2009) are the technical disciplines most fundamentally interrogated and put to test when addressing these topics.

Any systems compounding the FoS have to perceive a benefit of being part of the federation to engage in cooperation. Cooperating with other systems comes at a fixed, capital cost (implementing the capabilities to interface) and a recurrent, operational cost (every time a particular transaction is performed) which needs to be justified for the degree of additional benefit perceived by a system. From the perspective of a particular individual system and assuming rational decision-makers, the benefit obtained through cooperation in the federation has to come at a smaller cost than the cost associated to obtain the same benefit in a standalone fashion, in order to justify the transition.

Federations are still in their infancy in terms of our understanding of them, the available methods to analyze their emergence and evolution, and the methods to architect and design systems to operate in a federation. This thesis attempts to bridge some of these gaps for the first time. The next chapter specifies the thesis objectives and decomposes them in a set of fundamental research questions, as a way to guide the research efforts.

The methods needed to answer the research questions about FoS are to be drawn essentially from systems architecting theory, multi-objective optimization and tradespace exploration, utility theory, network flow problems, graphs and decision trees. The potential of several techniques found in the tradespace exploration literature (Ross and Hastings, 2005) in relation to FoS is discussed in the literature survey (Chapter 4) and approach (Chapter 5) sections of this thesis. Being FoS a new class of SoS, the shortcomings of some of the techniques and their applicability are also given special attention to. Through exploring, adopting and tailoring a suitable set of methods this work aims to build a cohesive framework to guide systems architects faced with the challenge of deploying a new system in federated environment. Additionally, the characterization of FoS needs of new theoretical constructs which are discussed in chapter 5.

The rest of this document describes the specific objectives, the approach and the expected results of the research proposed here. As mentioned, Chapter 3 outlines the thesis objectives and the related research questions. Chapter 4 introduces the necessary literature context related to both the problem and the methods envisioned to address it. Chapter 5 discusses the approach in detail, and describes all of the elements of the framework to answer to the research questions. Chapters 6,7 and 8 introduce a set of case studies, to be assessed in the light of the framework. The application of the framework to case studies responds to three needs. First and foremost, to exemplify and discuss specific implementation problems, second, to provide insights for the cases under examination, and third to provide validation and illustration of the chosen approach in more than one application.

Chapter 3 Thesis objectives

The problem of architecting FoS as introduced in the previous chapter mostly lies at the interfaces *between* systems. Traditional systems engineering is concerned with interfaces *within* systems. Nonetheless, influences coming from beyond system boundaries –exogenous– are also cause of concern of systems engineering and architecting disciplines, for which several approaches focusing on different problems (Aliakbargolkar, 2012; Ross, 2006) exist. However, the nature of the problem changes when these exogenous influences are indeed other systems subject to systems architecting efforts, and couplings between independent system architectures occur. The constituent systems in an FoS are not designed by a single architect or even in the same timeframe, and in many cases there might not be a single stakeholder with design authority over the whole set of systems. However, this does not mean we cannot influence or configure it through the interfaces and design of specific constituents.

As already outlined in the previous chapter, the general objective of the research proposed here is to provide *systems architects* with a cohesive approach to characterize engineering FoS, predict the evolution of an engineering FoS, and design their system taking into account the effects exerted from the federation, and to the federation, to/from their particular system of interest.

3.1 Research questions

The rationale for a particular system to operate within a federation resides in the cost and utility benefits achievable by cooperation, which must intuitively beat those achievable by standalone means for the cooperation scheme to be adopted. This includes the role of incentives between systems. In this work, we call synergy the benefits of cooperation between engineering systems. The concept of synergy is an essential theoretical foundation for this work, and will be further elaborated in Chapter 5.

This work needs to identify the conditions where synergy occurs as an emergence phenomenon in federations. Furthermore, quantifying synergy is the first step to support quantitative decision-making in early architecting processes. Through the use of this concept, we can formulate very specific research questions, which ultimately steer the construction of an architecting framework to give insights which are specific to FoS. These research questions are:

Q1) How can we measure synergy between a set of engineering systems?

A. What is synergy in engineering systems?
B. Under what conditions does synergy appear between systems?

Q2) How can we predict the formation and evolution of federation of engineering systems?

C. What are the effects of federating on a particular system?

D. How can we influence the long-term evolution of a federation?

The first research question is of a conceptual and exploratory nature, addressing the rationale for federations to exist, and attempts at quantification of the benefits of federating. From Q1, two sub-questions surface, aimed at thoroughly defining the aspects of synergy.

The second research question is targeted at the temporal aspect of federations which is very specific to this type of SoS. Due the managerial, operational and goals independence, systems will naturally join and abandon a federation at different times. This chronological element needs to be captured in any effort to analyze FoS. This question includes two sub-questions related to the systems architecting effort: how can we influence the evolution of an FoS, and how do we account for this when architecting a particular system?

3.2 Hypothesis

Based upon previous work (Lluch and Golkar, 2015), we can discuss in general what are the expected answers to the research questions of this thesis.

H1) In relation to the first research question, we expect that it is indeed possible to define and quantify synergy *in cost units*, reasoning through architectural tradespaces. Moreover, that synergy among engineering systems can occur on realistic scenarios;

H2) In relation to the second question, we expect that we can predict the evolution of a federation and understand the potential for wide adoption or decay of a specific federation; and that stable federations can exist under conditions to be investigated by this work.

3.3 Scope of the research

This thesis develops a set of methods for FoS applicable to the early phases of system conception. Table 2 shows a lifecycle model for an engineering FoS and the corresponding stage of *exploratory research* to which the methods developed here apply to.

Table 2. Generic lifecycle view, adapted from (Walden et al., 2015).

Life-Cycle stages	**Purpose**
Exploratory Research	Identify stakeholder's needs Explore ideas and technologies
Concept	Refine stakeholder's needs Explore feasible concepts Propose viable solutions
Development	Refine system requirements Create solution description Build system Verify and validate System
Production	Produce systems Inspect and verify
Utilization	Operate system to satisfy user's needs
Support	Provide sustained system capability
Retirement	Store, archive, or dispose of the system

This thesis is concerned with design and architecting from a strategic perspective, that is, taking into account delivery of value to stakeholders as an integrated result of system lifecycle. Questions of tactical nature, that is, how to operate a system within a federation or a coalition in an optimal manner, in a specific moment in time, fall into gaming approaches (Von Neumann and Morgenstern, 2007), which are out of scope of the present work.

Chapter 4 Literature review

This chapter discusses the literature which frames the research problem and introduces different tools required to answer our research questions. First, the background theoretical literature for FoS is introduced, which includes systems of systems (SoS) literature, and we add a note about the concepts of emergence and synergy. Then we discuss the systems architecting discipline, the field this thesis belongs to, together with its basic methodologies. In addition to the basic toolset, we present several advanced methods and their applications to illustrate how other authors tackle similar problems in other fields.

Finally, we close this review with practical considerations about satellite federations and their technologies, to familiarize the reader with the first case study of the thesis. The review here not only provides the reader with the necessary background from the existing work, but critically reviews and summarizes the latter, introducing the author's perspective on the SoS, Distributed Space Systems (DSS) and the shortcomings and advantages of several systems architecting methods.

4.1 Problem theoretical background

Federations of Systems are of type of SoS. As such, we need to position this thesis' research in the SoS literature, and make the necessary connections with overarching topics such as federalism, and functional emergence in systems.

4.1.1 Systems of Systems

While the term system(s) of systems had been used before, the first cornerstone work for SoS dates from 1998 when Maier provided a comprehensive definition and examined its elements (Maier, 1998). Two years before, Maier had already drafted an early version of this paper for the INCOSE international symposium (Maier, 1996). The early work of Maier listed five fundamental aspects distinguishing a SoS from a conventional system, which were *Operational independence of the systems composing the SoS, Managerial independence of the systems, Evolutionary development of the SoS, Emergent behavior, and Geographic distribution.*

The last aspect, *Geographic distribution*, concerns SoS which are based upon physically detached components, and also distributed in nature, such as computer networks. However, this quality alone is not substantial enough to justify a fundamental taxonomic change. Regardless the amount of added complexity distribution may add, on pure geographical distribution terms we are unable to distinguish a distributed system (Shaw et al., 1999) from a SoS. Therefore, we must look and resort to more fundamental differences between systems and SoS. From 1996 to 1998 Maier dropped this aspect from his definition set, rightly looking for more distinguishing aspects.

Emergent behavior concerns the surfacing of functions and behaviors which did not reside in any of the constituent systems of the SoS. In conventional systems, the term *emergence,* as will be discussed shortly, attempts to describe the new –sometimes unexpected– functionalities obtained when assembling a set of components, or subsystems. Since this taxonomy of elements, subsystems, systems and systems of systems is only a matter of perspective when decomposing a system, emergent behavior does not look neither as a particularly defining term for SoS. Notably, Maier also dropped it from 1996 to 1998.

The Evolutionary development of the SoS aspect was also unsurprisingly abandoned as an SoS characteristic on Maier's 1998 paper. While it is clear that *evolvability* is a key issue for SoS, as they are a portfolio of systems, there is a rich literature on system's evolution analysis. Evolution is not a feature unique to SoS.

Then, what can we use to define an SoS? By 1998, Maier had kept only two of the aspects remaining from his original attempt, which are operational and managerial independence. *Operational independence* means that if the component systems are to abandon the SoS, they can still operate for a useful purpose. *Managerial independence* is related to independent acquisition, ownership and governance of a particular system. These two aspects alone succeed to establish a clear taxonomic distinction, since in conventional systems these notions of operational and managerial independence of the systems´ components do not apply.

Based upon these observations, Maier (1998) further distinguished sub-types of SoS, called *directed, collaborative, and virtual,* based upon the degree of managerial and operational independence present. Table 3 illustrates the SETI@home project for extraterrestrial live search, a collaborative SoS.

Table 3. Illustration of SoS features as per Maier 1996 on the SETI@home project (Anderson et al., 2002).

	The SETI@home project
Geographic distribution	Using computers of volunteers all around the world
Emergent behavior	The processing of large datasets which is not achievable by single desktop computers alone
Evolutionary development	Different computers can join or leave, the software for data processing is patched and upgraded
Operational independence	The computers retain the ability to perform the original tasks they were acquired for. The SETI data sets are separated so that de-coupled processing is possible.
Managerial independence	The computers belong to different individuals and institutions

Directed SoS are those where component systems are built and managed for specific purposes, but retain a degree of operational and managerial independence, such as integrated air defence networks. Military SoS generally belong to this subtype.

In collaborative SoS, there is no centralized authority to operate the SoS, yet the systems associate to fulfil some higher goal. This would be the case of the SETI@home project (Anderson et al., 2002) where users lend their computing power to process the project's datasets. In contrast, in a virtual SoS, there is not even a higher goal agreed among the systems. That is, the systems *cooperate* rather than *collaborate* to fulfil their own local purposes.

Maier's principles of managerial and operational independence constitute the most accepted understanding of what a SoS is (Mekdeci et al., 2014), notwithstanding other definition attempts (Jamshidi, 2009) which account for a significant part of SoS literature. With noteworthy exceptions, the additional definitions are not as satisfying as Maier's principles. Besides Maier's, another widely used definition is the one issued by the US Department of Defense (DoD):

"A SoS is a set or arrangement of systems that results when independent and useful systems are integrated together into a larger system that delivers unique capabilities" (DoD, 2008).

DoD is one of the organizations worldwide with the most practical experience in SoS, and their definition is a working one tailored to their mission. '*Independent and useful systems'* resonates with the independency and managerial independence of Maier. However, the delivery of unique capabilities remains unclear as conventional systems are also built to deliver unique capabilities. Indeed, that is similar to the emergent behavior aspect discussed by Maier, which is beyond doubt a cornerstone to SoS but of modest taxonomical value.

Other authors added additional relevant aspects to the definition. Boardman and Sauser (2006) discuss SoS on the dimensions of autonomy, belonging, connectivity, diversity, and emergence. As we discussed, dimensions such as emergence, connectivity and autonomy are of uttermost importance in SoS, but they are only quantitatively different than in a conventional system. The aspect of belonging, however, connects with federalism's principles as we discuss in the next section: on the federalism lexicon, a *dual citizenship* exists when a system which has its own goals but also chooses to participate in an SoS. The diversity dimension, while not unique, is also exacerbated in SoS, where constituent systems may be heterogeneous in function and forms. De Laurentis (2005) highlighted the SoS control -related to Maier´s principles- and connectivity as fundamental aspects of an SoS, and went on using this taxonomy to guide the *design* methods for SoS. A fundamental conclusion of that work is that each SoS typology will require specific methodologies, which is a vision shared by this thesis.

The abundance of literature on SoS at the conceptual and definitions level has led to a certain degree of confusion and misclassification on the field. While some authors strongly emphasize complexity (Jamshidi, 2009), and SoS are indubitably complex entities due the additional level of interfaces, complexity alone makes SoS indistinguishable from Complex Adaptive Systems (CAS) which have their own literature. Adding more decomposition levels to traditional systems design methods has also been proposed, but does not do complete justice to the SoS field. Finally, some unfortunate usage of the SoS term in the literature as a catch-all, universal concept (encompassing from galaxies to rocket engines) has led to justified criticism from the systems engineering community.

Moreover, some authors have also engaged in the discussion about the need of a specific *SoS systems engineering* field. Such a discussion is out of scope of the proposed research, however a very informative and practical view of SoS systems engineering implementation is given by Dahmann in (Dahmann et al., 2011).

This thesis adopts systems architecting as the central discipline from which to draw concepts and methodologies to apply also to SoS. In this work, SoS are a very specific entity, strictly defined by operational and managerial independence. Furthermore, we will narrow down the scope to what Maier called *virtual* System of Systems, the maximum expression of managerial and operational independence. Maier himself used the expression *´federated´* for this type of systems (Maier, 1998). Hence, the next section introduces the work of Sage and Cuppan (2001), who specifically addressed federalism in engineering systems. Table 4 summarizes the SoS taxonomy presented here.

Table 4. Proposed SoS taxonomy based on the sources of this chapter.

	Directed SoS	**Collaborative SoS**	**Virtual SoS /FoS**
Operational independence	Operations centralized when pursuing SoS goals	Permanently independent	Permanently independent
Managerial independence	Management centralized when pursuing SoS goals	Permanently independent	Permanently independent
Central purpose?	Global SoS goals exist	Global SoS goals exist	No global SoS goals exist
Systems interaction scheme	Hierarchical structure	Voluntarily Collaborate	Voluntarily Cooperate
Global Stakeholders?	Yes	Not always	No
Examples	Military integrated air defense systems (Maier, 1998), robot swarms (Jamshidi, 2009), smartgrids	SETI project (Anderson et al., 2002), VSMs (Matevosyan et al., 2015), GEOSS (Lautenbacher, 2006)	FSS

4.1.2 Federalism principles

The principles of federalism come from political theory and state organization. Yet, as a governance method, it can be applied to corporate governance or to engineering systems management. Handy's principles of federalism for organizations were published in 1992 in the Harvard Business Review (Handy, 1992) and adopted by Sage and Cuppan to dissert about SoS, families and federations of systems.

These principles of federalism are *Subsidiarity, Interdependence, A uniform way to do business, Separation of powers and Dual citizenship.*

Subsidiarity means that decision-making remains in the lower, component-level parts instead of a centralized overseeing authority. *Interdependence* means that systems interact because they need each other and agree to freely combine (and/or centralize) tasks whenever necessary, but based on mutual autonomy and anytime revocability by the participants. *A uniform and standardized way to do business* is related to having common communication procedures and rules of conduct. For engineering systems this has practical implications on interfaces and communications protocol. *Separation of powers* means that daily operations, monitoring and long-term governance of an FoS are to be carried out by different bodies. This concept is less applicable than others to engineering FoS. *Dual citizenship* means that each system is a 'citizen' of two communities, their local one and the federation. Systems need to have a purpose and identity by themselves, while simultaneously belonging and having a stake in the overall FoS.

Maier´s virtual SoS concept and operational/managerial independence principles are useful for understanding *'what'* is a federation. The federalism principles above capture instead the *'how'* and *'why'* of an FoS. The *'why'* is ultimately related to mutual benefit, or synergy.

4.1.3 About emergence, collaboration, and cooperation

The concept of emergence is ubiquitous in systems engineering. As Crawley defined, *"Emergence refers to what appears, materializes of surfaces when a system operates. [...] Most obviously and crucially, function emerges"* (Crawley et al., 2016a). Systems are an assembly of different components to generate a new function of some kind, which is a desired emergence that responds to the design intent. For instance, putting concrete and steel together in an organized fashion for a bridge allows safe passage over a river. The thousands of components of a car enable personal transportation. However, other types of emergence are not always anticipated or even desirable, such as catastrophic bridge collapses or car accidents.

Another way to think about emergence is as '*the whole is more than the sum of the parts*' as enunciated by Aristotle. The systems architecting discipline then aims at anticipating and managing emergence, while avoiding potentially damaging outcomes of it. Emergence is also a key feature to understand SoS and specifically federations, as we discussed in the previous section, however it is not unique to it, but a general phenomenon of all systems.

As emergence is understood in terms of surfacing new functionality, emergence is by definition strongly connected to directed and collaborative SoS. On these subtypes of SoS, systems collaborate for a higher goal, what DoD calls '*deliver unique capabilities*'. This is indeed aligned with the general linguistic meaning of collaboration, *'to work with another person or group in order to achieve or do something'*. Collaboration includes a notion of achieving a higher purpose that was not achievable by participating entities alone. In the case of a pure virtual SoS, or FoS, there is no higher goal, yet the systems participate to support they own local goals, using their own measures of cost-benefit. Table 4 captures this distinction between SoS.

Hence, the emergence focus on FoS is not on new functionalities –notwithstanding them– but on enhancing local system functionality present *a priori*. This type of relation is better worded as cooperation. Indeed, the Merriam-Webster dictionary definition for cooperation *is "the increased effectiveness that results when two or more people or businesses work together"*. Translated to our discussion on the engineering systems domain, increased effectiveness is related to supporting native system(s) functionality instead of new functionality. The result of cooperation is then some form of emergence, but instead of the primary one related to new functions, it is related to enhancing functions on the component systems. In this work, we call this result of cooperation *synergy.*

4.1.4 Synergy in the literature

Synergy is a term mostly unused in engineering research. It is instead present in corporate management and industrial economics research (Chatterjee, 1986), and in the biotechnology and computational biology fields (Griffith and Koch, 2014). The latter is concerned by coalitions of random variables $\boldsymbol{X}$ trying to cooperatively predict the value of a target random variable $\boldsymbol{Y}$, to model phenomena such as neuronal interactions and combination of genetic information. The former is concerned about synergies within and between businesses.

Evaluating synergies between businesses is of utmost importance when considering corporation mergers. Chatterjee (1986) claims that, in a corporate merger, the increase of corporate value derives from the opportunity created by the resulting coalition to utilize a scarce resource. The return of the merger then depends on the actual resource scarceness, the effectivity of the merger implementation, and the availability of market opportunities for the resource. Discussing further the nature of merger synergy, Chatterjee distinguishes synergy types in mergers in three categories: collusive, operational and financial synergy.

Collusive synergy refers to the advantages on market power of joining businesses. Operational synergy is internal to the merged company and refers to production or administrative cost advantages. Finally, financial synergy refers to the improved access to the capital market for a bigger company. Chatterjee goes on into evaluating the effect of synergy in mergers, using the Cumulative Abnormal Returns (CAR) measured on company stock prices to analyze empirically the value created in mergers.

While the features of operational synergy could be arguably identified in FoS, the definitions of collusive and financial synergy do not readily apply to engineering systems. In a more general approach, including mergers, diversifications, and acquisitions, Iversen (1997) studied the nature of corporate synergy building upon several works (Ansoff, 1987; Markides and Williamson, 1994; Penrose, 1995). Iversen lists from the literature a wide scope of synergies at product, organizational and financial level.

In a brief but powerful observation, Iversen emphasizes that there are synergies not explained purely by economies of scope, by discussing in terms of Return on Investment (RoI); as *RoI(a+b)>RoI(a)+RoI(b).* That is, RoI can behave as a *superadditive* function when combining two assets, resources or companies. Additionally, as classifying elements, Iversen distinguishes a *static* synergy category, related to depictions of a company as a collection of fixed assets, a *dynamic* dimension, related to company competences, the aforementioned *superadditive* component, and also *subadditive* component. Crossing these categories we obtain a 4x4 matrix, which Iversen populated with the different synergy types identified in the economic literature.

Table 5 captures Iversen's taxonomy for corporate synergy. Subadditive synergies are related to sharing asset amortization, while superadditive synergies are about creating a unique scarce resource, to paraphrase Chatterjee (1986). In engineering systems, we expect to see examples of both. The discussion here established our departing point to define and quantify the concept of synergy for engineering systems, introduced in detail in Chapter 5.

Table 5. A Taxonomy on synergy in corporate economics, adapted from Iversen (1997).

	Static	**Dynamic**
Subadditive	Economies of scope /asset amortization	Competence amortization
Superadditive	Sales, operating, investment and managerial synergy (Ansoff, 1987)	Complementarity

4.2 On the systems architecting discipline

A system architecture is *"an abstract description of the entities of a system and the relationships between those entities*" (Crawley et al., 2004). Systems architecting as a discipline attempts to guide the very early decisions of devising a *complex* system. To do so, it relies on a set of heuristics and conceptual tools. Rather than attempting to find optimal design solutions, the system architecting process reflects from a clean sheet perspective on the relations between functions and system embodiment (the form), the interfaces between system components, on what are the main decisions and fundamental issues to address in the design process, and what are the design influences (specifically stakeholder needs) that have to be accounted for.

Systems engineering (NASA, 2007) is an older discipline, existing in academia for more than 50 years, but in practice much older. Systems engineering addresses the full lifecycle management of engineering products. Systems engineering *"is an iterative process of top-down synthesis, development, and operation of a real-world system that satisfies, in near-optimal manner, a full range of requirements for the system*" (Eisner, 2008). Systems engineering principles are routinely applied to multi-disciplinary problems in spacecraft and airplane design, software development, robotics, and many other fields.

Systems engineering is a holistic thinking frame that recognizes the importance of managing interfaces between subsystem components. The value of systems engineering practices in real projects includes cost, risk and schedule improvements (Walden et al., 2015).

Systems engineering includes all phases of design, operations, support and decommission of the system under study. There are a number of published standards giving a perspective on system lifecycle phases ("ISO/IEC/IEEE International Standard – Systems and software engineering – System life cycle processes," 2015) and several organizations have their own standards (DoD, 2008; NASA, 2007) . A generic lifecycle was illustrated in the previous chapter, section 3.3.

There has not been much documented discussion about the exact relation between systems engineering and architecting disciplines, or if one encompasses the other. The most intuitive conception is that architecting as a process is performed on the very early conceptual phases of a system development, in the exploratory and/or concept stages. The systems architecting discipline is fundamentally a response to complexity (Crawley et al., 2016a).

This thesis belongs to the field of systems architecting and is intended to support decision-making in the exploratory research and conceptual phases of system lifecycle. One of the fundamental quantitative tools of systems architecting is tradespace exploration, which is a cornerstone method of this thesis and is introduced next.

4.2.1 The tradespace exploration paradigm

Tradespace exploration is a quantitative systems architecture comparison method, not concerned about detailed design, but by understanding how different architectural alternatives fulfil the needs of system stakeholders (Ross and Hastings, 2005).

A *tradespace* is a set of different alternative system architectures compared by using a set of metrics. The goal of tradespace exploration is therefore to characterize the space of alternative architectures for a given system and quantify them with metrics relevant to the stakeholder. This quantification, based on an end-to-end system model, supports decision-making as it highlights which architectures *dominate* others on the set of metrics chosen. The tradespace exploration method can be divided in six steps illustrated in Figure 3.

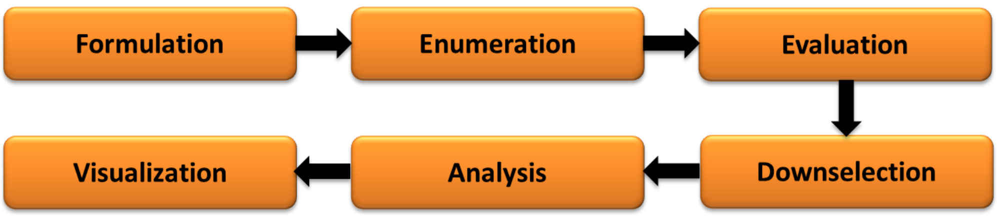

Figure 3. The steps of tradespace exploration.

In the formulation step, one chooses the relevant set of *decision variables* $\boldsymbol{X}$, which acts as input to the system model. Those represent variables that can be controlled by the architect.

In the enumeration step, all the combinations of potential values of decision variables are listed. Each combination is a different instance of the *design vector.* Usually decision variables are quantified in discrete steps as shown in Table 6, leading to a finite (although sometimes very large) set of design alternatives. For instance, in the case Table 6 shows, 5*4*3=60 different designs are possible. When all of the possible alternatives are assessed, we are doing *full factorial* tradespace exploration. When this is not possible, one must resort to Design of Experiments (DOE) techniques to partially explore the combinatorial of decision variables (Uy and Telford, 2009).

Table 6. Notional example of decision variables and the values to be explored.

Decision Variables	*Possible values*
Number of satellites	{2,4,6,8,10} (5 possibilities)
Instrument aperture [m]	{0.2, 0.5,1,2} (4 possibilities)
Orbital altitude [km]	{400,600,800} (3 possibilities)

In the evaluation step, we feed the design vectors into an end-to-end system model. The goal of this exercise is to assess each alternative architecture with a set of metrics. Hence, the system model we build must be able to output the metrics of interest. Evaluation metrics can be simple proxies, or elaborate combinations of utility functions if accurate descriptions of stakeholder needs are available.

In the downselection and analysis steps, we represent the different architectures in the space of the chosen metrics. A set of promising architectures can then be selected for further analysis, or a particular region of the tradespace re-assessed via refining the discretization of the decision variables.

The visualization step is dependent on how many metric dimensions we want to assess. 2D representations of a cost and performance metric are common (Palermo et al., 2015). Plotting tradespaces with three or more simultaneous metrics is challenging and is less common, as complex representations somewhat beat the purpose of decision-making support. Figure 4 illustrates a notional tradespace of 20 architectures for a space earth observation spacecraft, represented by discrete points. The metrics chosen are achievable sensing resolution against system development and launch cost.

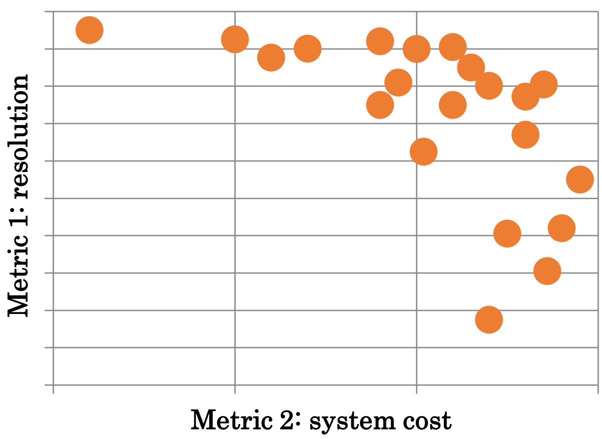

Figure 4. Notional tradespace depiction for an Earth observation spacecraft: Cost and resolution performance metrics for 20 architectures based on different architectural decisions.

It is essential to select adequate metrics when assessing architecture alternatives. In order to capture insightful trades, the metrics must be conflicting. Tradespace exploration does not yield single optimal design points, nor is this possible at early stages of system lifecycle when the requirements are not solidly grounded. Its intent is instead to capture the relation between fundamental metrics, and to narrow down the design search space by identifying *dominant architecture sets.* By using the Pareto-optimality principle (Chinchuluun, 2008), it is possible to screen for superior architectural alternatives, as discussed in the next section.

4.2.2 Multi-objective optimization and Pareto optimality

Multi-objective Optimization (MOO) (Andersson, 2000; De Weck, 2004; Marler and Arora, 2004), deals with finding a vector of independent variables $\boldsymbol{X}$ that minimizes a non-singleton set of objective functions $\boldsymbol{J}$. Formally, this is described as:

$$\begin{aligned}
&\min \boldsymbol{J}(\mathbf{x},\mathbf{p}) \ \ where\ \boldsymbol{J} = [J_1(\boldsymbol{x}), J_2(\boldsymbol{x}) \dots]^T \\
&s.t.\ \boldsymbol{g}(\boldsymbol{x},\boldsymbol{p}) < 0 \qquad \boldsymbol{x} = [x_1, x_2, x_3 \dots]^T \\
&\boldsymbol{h}(\boldsymbol{x},\boldsymbol{p}) = 0 \qquad \boldsymbol{g} = [g_1, g_2, g_3, \dots]^T \\
&x_{i,lb} \le x \le x_{i,ub} \quad \boldsymbol{h} = [h_1, h_2, h_3, \dots]^T \\
&\boldsymbol{x} \in S
\end{aligned} \qquad \text{Eq. 4-1}$$

$\boldsymbol{J}$ is a column vector of objective functions J_i. These objectives are dependent on a set of independent variables x_i, which are subject to equality constraints $\boldsymbol{h}$ and inequality constraints $\boldsymbol{g}$. Furthermore, the independent variables x may be bounded from below (x_{lb}) or top (x_{ub}). A solution for $\boldsymbol{x}$ in the feasible space S respects the constraints $\boldsymbol{g},\boldsymbol{h}$, and the bounds. Through this document, optimization will be exemplified as a minimization problem, note this is equivalent of a maximization changing the sign of the objective function(s) (Bertsimas and Tsitsiklis, 1997).

MOO addresses real life problems from managerial and systems engineering perspective. Project management goals are effectively multi-objective as risk, cost and schedule need to be minimized while product performance is met. Systems engineering problems are also well expressed in terms of MOO. For instance, for an earth observation spacecraft we may want to maximize sensing performance in resolution and coverage, while minimizing platform mass, manufacturability, development cost, or others.

Even without attempting a rigorous optimization on specific target functions, it is to the advantage of design efforts to capture and visualize the conflicting trades between system metrics, as discussed before in the tradespace exploration section. Indeed, tradespace exploration combined with the *Pareto dominance* principle is one of the techniques of MOO, often categorized as *a posteriori preference method* (De Weck, 2004). In contrast, *a priori* methods to tackle MOO problems involve some sort of consolidation of the objective functions before attempting the optimization. Table 7 introduces a classification on methods to address MOO.

Table 7. A classification of MOO methods, adapted from (De Weck, 2004).

Scalarization methods **(a priori expression of preference between objectives)**	**Pareto methods** **(a posteriori expression of preference)**
Weighted Sum approach	Exploration and Pareto filtering
Compromise programming	Weighted sum approach (with weight scanning)
Multi-attribute utility analysis (Ross, 2006)	Adaptive weighted sum
Goal programming	Normal Boundary Intersection (NBI)
Lexicographic approaches	Multi-objective genetic algorithm
Acceptability functions & fuzzy logic	Multi-objective simulated annealing

From scalarization methods, the most relevant approach to the work here is multi-attribute utility analysis, which is fundamental to Multi-attribute Tradespace Exploration (MATE). Both topics are covered in the next two sections of this literature review. From Pareto methods, we proceed now to explain the basics, which is exploration and Pareto filtering. *Pareto dominance* is a concept drawn from economy theory (Pareto, 1906). Be $\boldsymbol{J^A}, \boldsymbol{J^B}$ two feasible objective vectors. Then $\boldsymbol{J^A}$ is said to *weakly dominate* $\boldsymbol{J^B}$ in a minimization assignment if:

$$\boldsymbol{J^A} \leq \boldsymbol{J^B}\ \forall \boldsymbol{i} \qquad \text{Eq. 4-2}$$

$$\text{and } \boldsymbol{J^A} < \boldsymbol{J^B} \text{ for at least one i}$$

For strong dominance, all elements in vector $\boldsymbol{J^A}$ need to be strictly smaller than the corresponding elements in $\boldsymbol{J^B}$. Pareto dominance is not a reflective property**,** that is, if $\boldsymbol{J^A}$ does not dominate $\boldsymbol{J^B}$, this does not mean $\boldsymbol{J^B}$ dominates $\boldsymbol{J^A}$. Therefore multiple, non-dominated solutions are possible. The set of non-dominated $\boldsymbol{J}$ vectors is called the *Pareto-optimal set.*

Exploration and Pareto filtering involves sampling the design space $\boldsymbol{x}$, as explained in the previous section, and then identifying the Pareto set. Besides the Pareto set, other particular objective vectors in the MOO problem receive specific names, such as Ideal, Nadir, or Utopian. For a full lexicon on these particular solutions, refer to (Narzisi, 2008). In terms of nomenclature, also note that the term *dominance* refers to the objective vector space and the corresponding term on independent variables space $\boldsymbol{x}$ is Pareto *efficiency* (Narzisi, 2008).

As an example, consider three sets of independent variables embodied by $\boldsymbol{x^A}$, $\boldsymbol{x^B}$, and $\boldsymbol{x^C}$, which belong to the feasible set S. For simplicity of this explanation, assume an optimization problem aimed at minimizing two objectives as in $\boldsymbol{J}= [J_1,J_2]$.

We can them sample the objective functions vector J for each set, yielding a different notional result for sets $\boldsymbol{x}^A$, $\boldsymbol{x}^B$, and $\boldsymbol{x}^C$, as:

$$\boldsymbol{J}^{\boldsymbol{A}} = [J_1(\boldsymbol{x}^{\boldsymbol{A}}), J_2(\boldsymbol{x}^{\boldsymbol{A}}) = [90{,}750]$$

$$\boldsymbol{J}^{\boldsymbol{B}} = [J_1(\boldsymbol{x}^{\boldsymbol{B}}), J_2(\boldsymbol{x}^{\boldsymbol{B}}) = [120{,}175] \qquad \text{Eq. 4-3}$$

$$\boldsymbol{J}^{\boldsymbol{C}} = [J_1(\boldsymbol{x}^{\boldsymbol{C}}), J_2(\boldsymbol{x}^{\boldsymbol{C}}) = [120{,}801]$$

We can say then that $\boldsymbol{J}^c$ is strongly dominated by $\boldsymbol{J}^A$ as the latter is smaller in both objectives. $\boldsymbol{J}^c$ is weakly dominated by $\boldsymbol{J}^B$, as they are equal on the first objective. Most interestingly, $\boldsymbol{J}^A$ and $\boldsymbol{J}^B$ do not dominate each other, and belong to the Pareto set in this example. Intuitively, two non-dominated objective vector solutions capture a tradeoff: either we pick the minimal solution in one objective, or on the other. These cases are picked from our example tradespace exploration of Figure 4. As $\boldsymbol{J}^A$ and $\boldsymbol{J}^B$ are not dominated by any other set, they are part of the Pareto set, together with the points highlighted Figure 5.

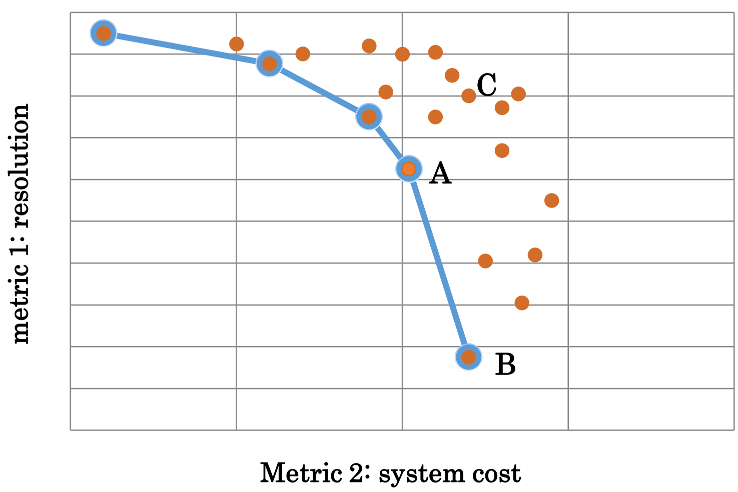

Figure 5. Tradespace exploration for Earth Observation spacecraft with Pareto front highlighted. Architecture A strongly dominates C. A and B do not dominate each other.

From the architectures sampled in Figure 5, five are non-dominated by any others and this constitute the Pareto-optimal set. When using such graphical representations, the term *Pareto front* or frontier is widely used to describe the Pareto-optimal set. Many real life problems lack an actual optimal on all objectives, instead having non-singleton Pareto sets composed of non-dominated solutions as shown above. Note that Figure 5 represents the architectures in a space of two metrics sampled discretely; therefore there are no assurances of the actual optimality of the non-dominated architectures if the independent variable set $\boldsymbol{x}$ is continuous. Furthermore, if the set $\boldsymbol{x}$ is composed of discrete variables, we can only assure that we have identified the actual Pareto set if we sample all the set potential values, which is often called *full factorial* exploration.

Optimality assurances for a MOO problem can be obtained by satisfying the generalized Karush-Kuhn-Tucker conditions (De Weck, 2004):

$$\boldsymbol{x}^* \in S \; and \; S = \emptyset$$

$$J, g, are \; differentiable$$

$$\boldsymbol{\lambda_j g_j}(\boldsymbol{x}^*) \leq 0 \; \forall \; j = 1,2, \dots m \qquad \text{Eq. 4-4}$$

$$\boldsymbol{\lambda_j g_j}(\boldsymbol{x}^*) = 0 \; where \; \boldsymbol{\lambda_j} \geq 0 \; \forall \; j = 1,2, \dots m$$

there exist $\mu_i \geq 0 \quad \forall \; i = 1,2, \dots n \; with \; at \; least \; one \; i \; such \; that$

$$\mu_i > 0$$

$$\sum_i^n \mu_i \nabla J_i(\boldsymbol{x}^*) + \sum_j^m \boldsymbol{\lambda_j} \nabla g_j(\boldsymbol{x}^*) = 0$$

When functions are differentiable, the results' optimality of the methods listed on Table 7 can be checked. Within the listed techniques requiring a priori understanding of preferences, multi-attribute utility is a widely used method. Its foundations are explored next.

4.2.3 Utility theory fundamentals

Expected utility theory (Arrow, 1966; Von Neumann and Morgenstern, 2007) has been widely used to capture stakeholder preferences and 'usefulness' of engineering designs. The notion of utility in engineering design is a narrow interpretation of the more ambiguous notion of *'value'* (Collopy, 2009). In engineering, a *value model* maps a collection of attributes to a numerical scale with the goal to support comparisons of attributes based on preference. In the scope of systems engineering design and architecting, expected utility theory can be understood as a specific value model to support rational decisions of a stakeholder based on expected outcomes. Hence, an expected utility of a decision can be computed as a weighted sum of the utilities of the potential outcomes of the decision, using probability of each outcome as weights. In systems design and architecting, a decision can be assimilated to adopting a specific design or architecture, or in our nomenclature, a set of decision variables $\boldsymbol{x}^*$. Utility is a measure of value with the following properties (Von Neumann and Morgenstern, 2007):

$$u \rightarrow p = \mathrm{v}(u) \qquad \text{Eq. 4-5}$$

$$for \; any \; two \; u \; and \; w, only \; one \; relation \; holds:$$

$$u = w, u > w, or \; w < u \qquad \text{Eq. 4-6}$$

$$Now, if \; u > w \; implies \; v(u) > v(w) \qquad \text{Eq. 4-7}$$

$$\mathrm{v}(\alpha u + (1-\alpha)w) = \alpha v(u) + (1-\alpha)v(w) \qquad \text{Eq. 4-8}$$

$$if \; u > w \; and \; w > x, imply \; u > x \qquad \text{Eq. 4-9}$$

In Eq. 4-5, the transformation assigns a number p to a utility u based on a utility function v. The condition in Eq. 4-6 implies that the preference system is complete. The condition in Eq. 4-7 represents the requirement to any utility function to maintain the preference of utility u to w. Next, Eq. 4-8 is what von Neumann and Morgenstern called 'a natural operation' which, if supported by the utility function system, allows proving that utilities are numbers p up to a positive linear transformation.

In other words, utility theory is based upon mapping preferences into a numerical system which preserves ordinality, within an arbitrary scale which can be then positively linearly transformed to any other without loss of preference information. Preference is also transitive, and the mapping of utility into numbers p needs to reflect this (Eq. 4-9).

4.2.3.1 Multi-attribute Utility Theory (MAUT)

Typically, the decision process that needs to be supported in systems engineering, and specifically when applying MOO, presents more than one attribute on which to assess utility. If we assume *mutual utility independence*, and the utility functions to be multiplicative, the expected utility of a multi-attribute decision problem looks like (Ross et al., 2010):

$$K \cdot U(\hat{X}) + 1 = \prod_i^n (K\, k_i \cdot U_i(X_i) + 1) \text{ where } K = -1 + \prod_i^n (K \cdot k_i + 1) \text{ and } \sum_i^n k_i = 1 \qquad \text{Eq. 4-10}$$

For two utility functions (De Weck, 2004) the expression above would simplify to:

$$U(\hat{X}) = k_1 \cdot U_1(X_1) + k_2 \cdot U_2(X_2) + (1 - k_1 - k_2) U_1(X_1) U_2(X_2) \qquad \text{Eq. 4-11}$$

Note that the multiplicative is not the only form possible for utility functions. A generalized forms discussion can be found in (Abbas, 2009). Each utility function U_i maps a value for an attribute X_i to a numerical value on a scale between 0 and 1. Utility functions can adopt any general shape, from monotonic increase or decrease to convex and concave (De Weck, 2004). The k_i multiplier acts as a weight for each utility function. The K factor is required after adding individual utilities U_i to normalize again the combined utility $U(\hat{X})$ to the 0-1 range.

There are many methods to weight stakeholder preferences between attributes when building a MAU model. The most common is the lottery method. It is based on understanding at which point will the stakeholder switch from preferring a certain attribute combination to another one which is subject to a probabilistic process. Collopy (Collopy, 2009) provides a numerical example for the lottery method.

4.2.3.2 Limitations of expected utility theory for systems architecting

The first drawback when applying MAUT to systems architecting is related to finding adequate individual utility functions to capture appropriately the relationship between attribute fulfilment and value delivery to the stakeholder. If an adequate function can be found there is still a trade between fidelity and simplicity; it might be that a non-linear combination of non-convex functions describes very well the stakeholder attributes preference, but poses a complex problem in terms of optimization and interpretation of results. Conversely, overly simplified functions can sometimes not lead to useful or realistic insights for the architect and the stakeholder. Likewise, the systems value is not always convertible to utility functions or there is no explicit information on what is the value to the stakeholder of a specific attribute –see unarticulated value discussions (Ross, 2006) –.

The normalization and dimensionless of MAUT results make it difficult to obtain quantitative information between disparity of attributes or results in terms of physical quantities that might be better understood by the stakeholder (cost in dollars, mass in kilograms..)

Assigning weights is also a typical problem with this type of approach, which requires close interaction with stakeholders. When dealing with many stakeholders and dozens of attributes, it can be close to impossible to capture the weights by using lottery or indifference curves methods. Despite its limitations, utility theory and MAUT remains a technique in the system's architect toolbox due to its simplicity and synthesis power, and is the basis of Multi-attribute Tradespace Exploration (MATE), which is explained next.

4.2.4 Dynamic Multi-Attribute Tradespace Exploration

The MATE method uses MAUT to consolidate several attributes and enable a standardized utility against cost tradespace exploration. MATE, proposed by Ross and Hastings (Ross and Hastings, 2005), has been extended to a comprehensive framework to address uncertainty and exogenous changes, called the Responsive Systems Comparison (RSC) (Ross et al., 2008a). Ross developed RSC by adding a dynamic dimension to MATE, namely *epoch-era* analysis (Ross and Rhodes, 2008). This allows applying MATE to changing contexts to assess and support the sustained delivery of value of a system in a changing context. Figure 6 illustrates the epoch-era analysis.

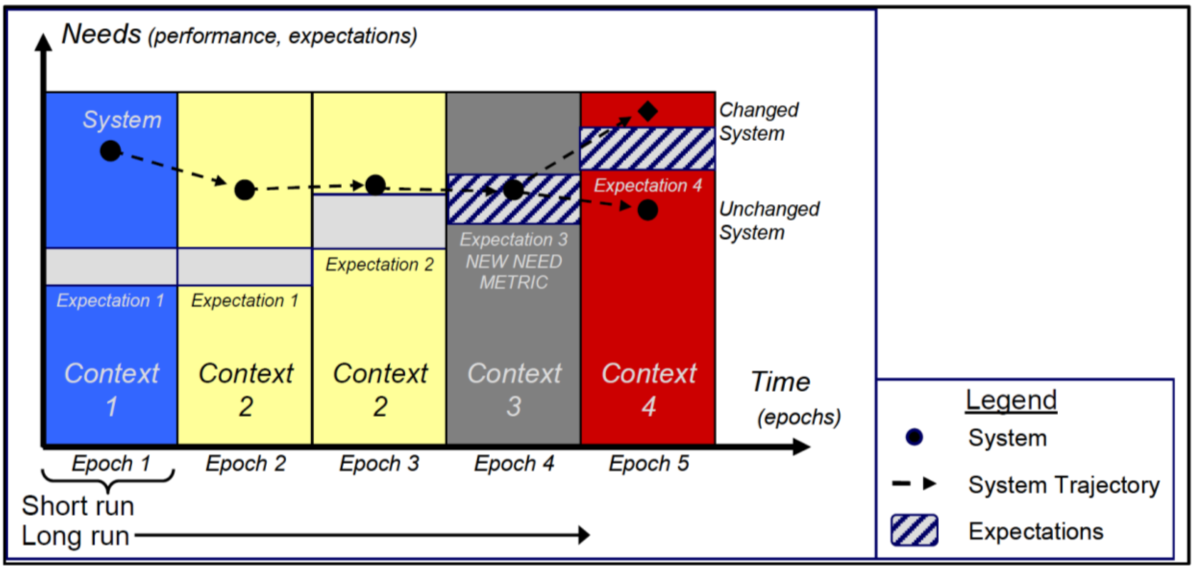

Figure 6. The changes on stakeholder expectations and operational contexts defines different epochs for tradespace exploration, from (Ross and Rhodes, 2008).

An *epoch* is a stage in time during the system's lifecycle where the architecting efforts are subject to static constraints, a fixed set of available technologies, feasible design concept and static articulated attributes. In order to introduce the dynamics of a potentially changing context for a system, Ross proposed to analyze the systems under a set of different epochs, which constitutes an *era.* Hence, the value delivery of a system can be quantified in a changing environment, which enables evaluating *system properties* such as robustness, flexibility, adaptability, and others, beyond what is possible with traditional static architecting techniques.

The RSC framework, with its main components MATE and epoch-era analysis, is relevant to this thesis both from a methodological perspective and in its application to SoS. However, at the end of this section we discuss the main differences and assumptions for FoS that will require a specific methodology. MATE and the RSC framework have been extensively applied to design for *ilities*, that is, system properties. The next section outlines the ilities field to illustrate the use of several methods of interest.

4.2.1 System properties: ilities

Advanced architecting methods, including MATE and epoch-era analysis, have been applied to the research of system properties, also called *ilities;* including flexibility, adaptability, robustness, and many others. The research on ilities is very illustrative of the application of various methods for capturing decision uncertainty and chains of decisions in systems architecting (decision analysis/decision trees, Markov chains, and advanced use of tradespace exploration) hence its interest for the research proposed hereby.

Ilities manifest themselves cross several engineering domains. Classic engineering ilities such as safety, reliability and quality have been studied for long (Adams, 2015), while others such as evolvability, adaptability or agility are gaining prominence as the complexity and value expectations of engineering systems rise.

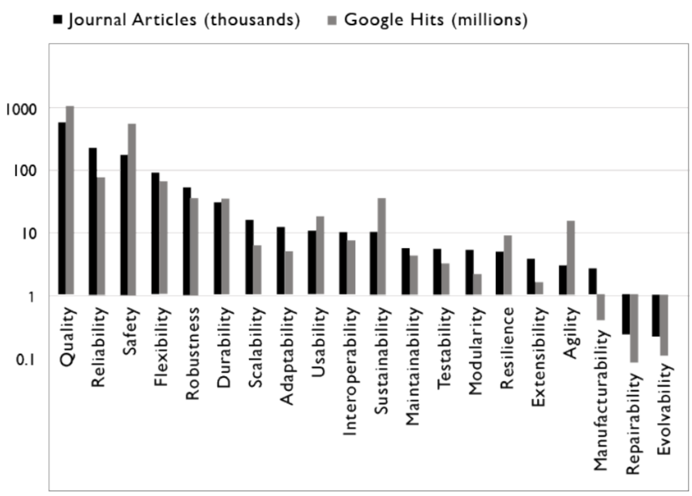

Figure 7. Occurrence of ilities in the literature and search engine hits, from (de Weck et al., 2012).

Figure 7 illustrates the occurrence of different ilities in the academic literature (de Weck et al., 2012). As shown, the academic literature has addressed tens of ilities, through definitions and evaluation methods which are, in general, very domain-dependent. Nevertheless, there are attempts at general definitions in the systems architecting literature, not only for each specific ility, but for the whole category. One of such definitions of ilities was given by de Weck and colleagues (2012), as *'desired system properties that often manifest themselves after a system has put to initial use'*. That is, for the most part, ilities are not part of a system's primary functional requirements, but can exert a lasting and wide influence after the system is deployed.

Ross and colleagues recently proposed to classify and define system ilities upon a semantic basis (Ross et al., 2011), reaching for linguistics theory to develop a rigorous, cohesive definitions space for ilities. It is useful to introduce the preliminary categorization they proposed, which distinguishes between change-related ilities, architecture related ilities, and new-ability-related ilities. Change-related ilities include, among many others, adaptability, agility, robustness and flexibility. Figure 8 introduces a comparative definition for these, adapted from Fricke and Shultz (2005).

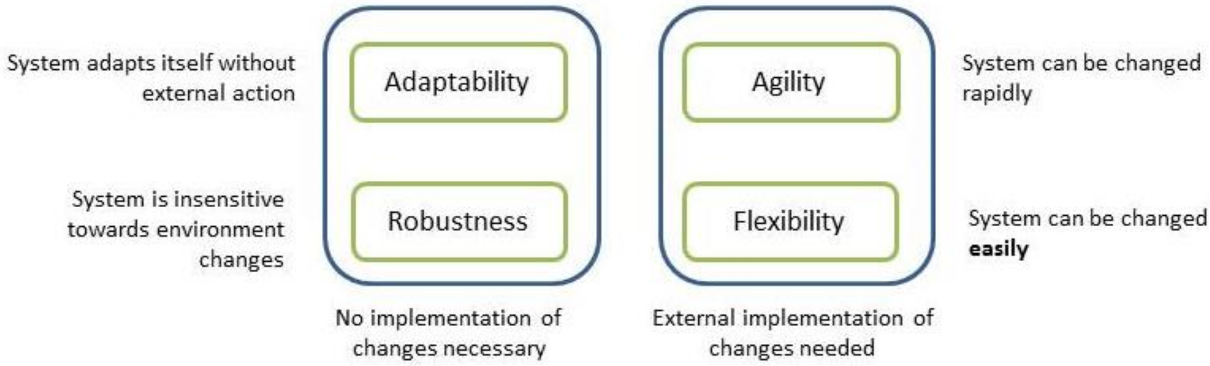

Figure 8. Four different aspects of changeability, adapted from (Fricke and Schulz, 2005).

Changeability is one of the key ilites which advanced architecting methods try to assess. Architectural ilities are a second class of ilities, including modularity, interoperability, independency, and others, which are in general enablers of other system ilities. While modularity may not be an interesting feature per se, it can effectively support upgradability and evolvability, as discussed later in the paper. A third class of ilities, ability-related, are those which deploy new capabilities in a system.

Now we introduce several methods to evaluate a few ilities on the design stages of a system. The methods presented herewith do not constitute a comprehensive literature review on the ilities subject, but a collection of relevant approaches and applications to the methods discussed in this thesis.

4.2.1.1 Flexibility

Flexibility is one of the most studied lilies. Several approaches exist to evaluate the flexibility, and two of them are presented here due to their illustration of systems architecting techniques.

4.2.1.1.1 Nilchiani's 6 elements framework

In her PhD, Nilchiani (2005) pays attention to system boundary, types uncertainty, system aspect to which flexibility is applied, time window of changes, response to change, and access to the system as the elements that configure flexibility. Nilchiani develops a multi-step framework leading to a representation of different alternatives in a delta-cost and delta-benefit tradespace. The fundamental part of this work is the presentation of the methods to assess the delta-cost and delta-benefit of each alternative respect a baseline. These methods include decision trees and real options analysis. Real options analysis allows evaluating the future value of an engineering or project option in the face of uncertainty. Nilchini draws from Saleh (2002) the application of Black-Scholes equations to aerospace applications, and also proposes binominal lattice and Montecarlo analysis in order to integrate the benefits and cost of all the possible ramifications of the decision chosen through the operation of an engineering system.

Nilchiani's framework excels at analyzing the value of embedding flexibility, that is, future options, to engineering systems, from a global project perspective; however it requires a design baseline for comparison, which already captures the stakeholder views.

4.2.1.1.2 Value-Weighted Filtered Outdegree

Viscito (2009) proposes a flexibility assessment integrated with Dynamic Multi Attribute Tradespace Exploration (Dynamic MATE). In dynamic MATE, the multi-objective performance of a set of alternative system designs is presented in a series of changing system contexts, called epochs.

Some designs may dominate others depending on the context. Ross and colleagues (2008b) introduced the concept of Pareto trace that accounts the number of contexts where a given design is Pareto-optimal. Furthermore, some design alternatives may allow to transition to a higher number of designs in chronologically subsequent contexts. This property was called the *outdegree* of a design alternative, and it is the gateway to flexibility. The outdegree of a design is limited by the amount of feasible transitions to other designs; therefore transition rules have to be established to characterize what are the acceptable transitions between design alternatives. The Value-Weighted Filtered Outdegree (VWFO) method adds a weight to the filtered outdegree to reward the designs with the potential to transition to high value designs in future contexts. This scheme evaluates flexibility as the potential to deliver value to the stakeholder across evolving contexts.

4.2.1.2 Scalability

Scalability plays a crucial role in any distributed system; consequently it has been analyzed extensively in communication and computer networks. Scalability is defined as the ability of the system to retain performance as some scaling variable –such as number of system components– increases. Hence, it is a desirable property of federated systems. Recent efforts by Portillo (2015) led to the development of a framework for the analysis of scalability in *Fractionated Satellite Networks*; a hybrid between FSS and Fractionated Spacecraft. Results show that the size limits of a federated system ultimately depend on the distribution of functionality between constituting elements.

4.2.1.3 Reconfigurability

Reconfigurability belongs to the domain of change abilities, and introduces an operations perspective: Reconfigurability is the ability to change system configurations "*repeatedly and reversibly"* (Siddiqi and de Weck, 2008) during different "*mission segments"* (Denhart et al., 2013). Among the methods to assess reconfigurability, Denhart's is compelling since it is based upon tradespace exploration and therefore a candidate for integration with other assessment techniques described before in this section. Denhart presents a reconfigurable system in a tradespace as a family of points, corresponding to different configurations of the system examined. By introducing the concept of 'surrogate points' Denhart is able to compare reconfigurable designs between them and to static, conventional designs, and establish the strong or weak dominance of certain system design alternatives.

4.3 Specific architecting efforts on SoS

A number of quantitative frameworks on *design* for SoS have been proposed, mostly devised for and applied to defence systems acquisition. As such, these SoS are *directed* and the term *design* can be, with caution, applied to them, even though they are composed of a portfolio of systems procured under different management and specific objectives. It is no accident that the existing frameworks address *directed SoS* since they are closer to conventional systems design and there is explicit interest by a stakeholder to have a decision support tool. Conversely, the lack of global perspective on collaborative and virtual SoS makes such a need for tools more elusive.

Davendralingam and DeLaurentis (2013) propose a robust framework for hierarchical, centrally controlled and clearly layered SoS. Their intent is to find an optimal SoS design through a range of perturbations, hence identifying robust designs. They define the SoS in net-centric perspective and set a mixed- integer linear program to maximize the set of overarching objectives to be carried out by interconnected nodes, or systems. The problem is subject to node connection capabilities and constraints. The robustness aspect is introduced through a slack variable, or conservatism parameter Γ that protects the linear constraints –which are uncertain– from falling into unfeasibility. Thus, the objective function is not excessively penalized by the uncertainty on the constraints and the obtained solutions are suitable for a range of constraints. This is known as the Bertsimas-Sim approach (Bertsimas and Sim, 2004). Said approach is then applied to the U.S. navy Littoral Combat Ship (LCS). The LCS is a modular system with requirements to perform in very diverse scenarios and equipment sets, but its classification as SoS is arguable. The proposed approach suits well the LCS and directed SoS, but cannot be readily applied to other SoS with lack of over-arching objectives and centralized design authority.

Sobieszczanski-Sobieski (2008) addressed SoS design by extending the integrated bi-level synthesis method to three levels. Although the decomposition of a system in n-levels is somewhat arbitrary and does not necessarily reveal the existence of an SoS, the methodology applied gives us a very much needed view of the SoS field from the perspective of multi-disciplinary design optimization (MDO). *Directed* SoS design is inherently subject to several engineering disciplines, a factor increasing the complexity of SoS modelling efforts.

Ricci and colleagues (2014) proposed the SoS architecting with ilities (SAI) method which is based upon MATE-RSC to identify paths and ility-enabling mechanisms for graceful evolution in SoS. The method starts with a very clear needs statement for the whole SoS by global stakeholders. From there it moves onto identifying potential system perturbations, which are operational context changes in line with the epoch definitions of the RSC method. Ricci's method then identifies desired ilities and ility-driving options. The classical steps of tradespace exploration then follow: formulation and enumeration of architectural alternatives, evaluation and analysis. Ricci exemplifies the steps of the method on a maritime security case study.

4.3.1 The SoS Tradespace Exploration Method (SoSTEM)

Chattopadhyay's work on architecting SoS (Chattopadhyay et al., 2008) also builds upon MATE and most notably, includes a detailed implementation of the proposed method. Chattopadhyay's thesis (Chattopadhyay, 2009) also includes a comprehensive review of SoS literature to which the interested reader is referred. Chattopadhyay addressed important aspects of SoS which are of practical architecting concern, such as control schemes, stakeholder influence and cost-benefit perception for participant systems. The distinction about the global perspective and local perspective is central to her work. SoS exist on a global perspective, have a mission and global stakeholders related to the former. Each system introduces also a local perspective with its own stakeholders.

In practical terms, her work is mostly applicable to directed SoS, but Chattopadhyay (2009) also explores some notions for virtual and collaborative systems through the assumption that participating in the SoS is optional for the component systems. To reflect on this idea, she introduces the concept of *perceived net benefit*, the rationale behind a system's participation in the SoS. Said participation leads to cost and benefit changes at the local perspective and also enables global benefits. These concepts are reasoned via an informal arithmetic: in order to join the SoS, the Perceived Net Benefit (PNB) for a local system must be larger than 0. The PNB depends on costs and benefits as shown in Eq. 4-12.

$$PNB = LB + GB - LC - GC + Inc \qquad \text{Eq. 4-12}$$

Where LB is the Local benefit, LC is the local cost, GB is the global benefit (linked to value delivery to an overall SoS stakeholder), GC is the global cost, and Inc are any Incentives in place to stimulate the System to join the SoS. Chattopadhyay indicates that a PNB marginally above 0 might not be enough to persuade systems to join the SoS, so a Threshold *Th* is introduced. Therefore, the condition for a given system to join a SoS would appear to be having a positive perceived net benefit above a certain threshold. In order to make the PNB larger than 0 including a threshold, incentives can be used. As Chattopadhyay points, when there is no incentive required, the systems might start collaborating spontaneously, generating a virtual SoS or FoS. This corresponds to a case this thesis defines as strong synergy, which is not the only way an FoS can occur, as explained in chapter 5.

As introduced before (Table 4), the taxonomical distinction between FoS and the rest of SoS types resides in the fact is that there is no explicit global value or stakeholders, notwithstanding the existence on non-stakeholder beneficiaries of the global FoS. Therefore the global terms in Eq. 4-12 vanish, since all benefits and costs are directly or indirectly accounted for in the component systems. However, for FoS architecting we need to have a deeper look at the incentives term. In the absence of centralized managerial control or global stakeholders, incentives in the sense Chattopadhyay (2009) uses them do not apply since there is nobody to distribute them. However, in FoS, systems can, peer-to-peer, indeed *incentivize* other systems to cooperate with them on the basis of a perceived benefit. In practice, this means that said systems trade some tangible or intangible commodity (resources, functions, services…) to achieve larger local benefit, on a commercial basis. The incentives term in Eq. 4-12 can then become positive or negative for systems depending if they receive or give the incentives. These concepts will be further developed in this thesis.

Chattopadhyay also discussed the effects of the absence of full managerial control, using the concept of influence: the ability of managers to persuade other systems to join the SoS. She studied the effects of managerial control and influence on the participation risk for systems in the SoS. These aspects play a role in the architecting framework developed on that work, the SoS tradespace exploration method (SoSTEM). The method makes use of consolidated SoS attributes, SoS cost models and a clear global SoS mission to compare SoS architectures on a single, consolidated global tradespace. Hence it is a adaptation of MATE on the modelling side to substitute conventional systems by SoS, and assess them on a similar utility-costs perspective, supporting decision-making by a global stakeholder. SoSTEM has been applied to multi-platform surveillance systems (Chattopadhyay et al., 2009). The value of the SoSTEM case studies is fundamentally notional but highly relevant since the SoS literature is unfortunately short of such detailed implementation exercises. Figure 9 gives an overview of the method. SoSTEM rightly includes legacy systems, new systems and their future evolution as design variables.

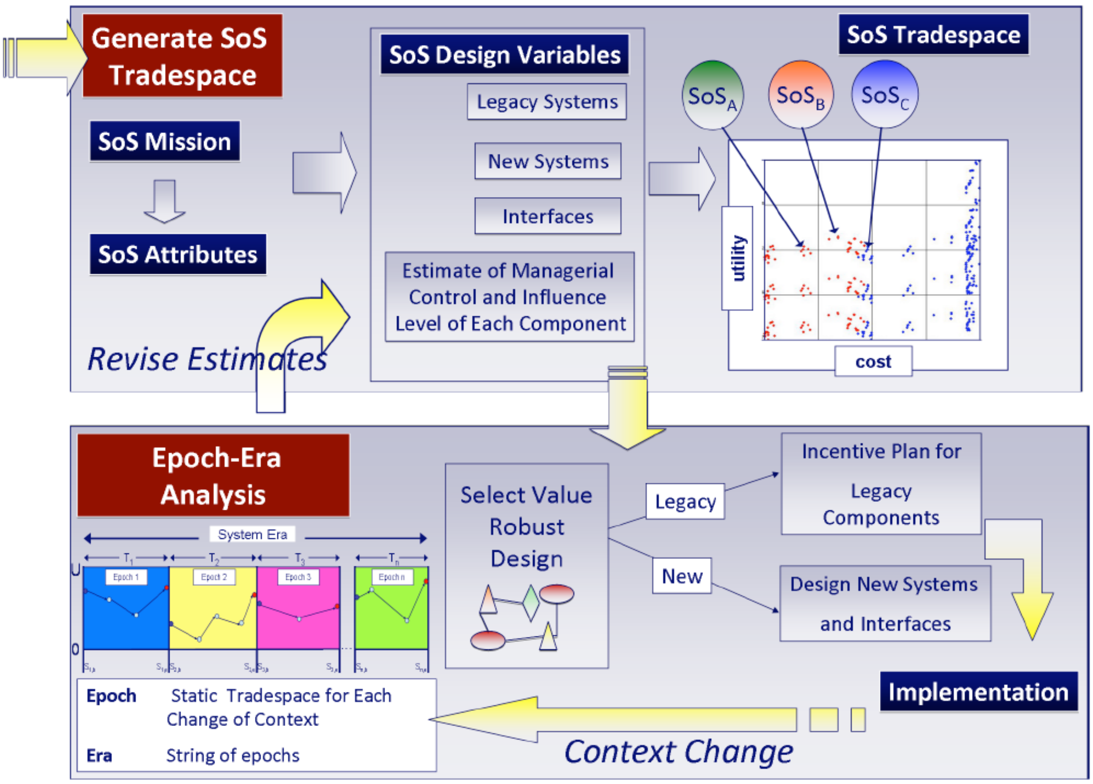

Figure 9. The SoSTEM method by Chattopadhyay (Chattopadhyay, 2009).

Despite these considerations, SoSTEM is mainly aimed at orchestrating single-step acquisition processes for the new systems as part of a SoS. This is very suited to directed SoS, especially in the defense domain, but does not bode well when applied to virtual SoS where there is no global perspective and no stakeholder to make decisions on a single SoS tradespace. For virtual and collaborative systems, we lack quantitative architecting frameworks. Yet, a set of heuristics and recommendations are available by Maier and Rechtin (2009).

4.3.2 Heuristics for collaborative SoS

Maier and Rechtin included in the *Art of Systems Architecting* book (Maier and Rechtin, 2009) a set of heuristics for collaborative SoS. They can be understood as principles for a practitioner attempting to influence the evolution of an SoS or managing its operations. The principles include *Stable intermediate forms, policy triage, leverage at the interfaces, and ensuing cooperation.*

Stable intermediate forms is introduced by the statement *"Complex systems will develop and evolve within an overall architecture much more rapidly if there are stable intermediate forms than if there are not"*. This notion originates in civil architecture, where it is a good practice for a building under construction to be self-supporting at several stages of its development. This reads as the need for technical, economical and politically self-supporting stages through the evolution of the collaborative SoS. This is especially important for FoS since it stresses the need for the federation agreement to yield tangible benefits as early as possible to the constituent systems, even when their numbers are small.

The *Policy Triage* is taken from medical emergencies field as a reflection on what is the scope of the design efforts on a collaborative SoS; what to control and what not. The collaboration scheme needs neither be overregulated- hindering emergent collaboration possibilities-, nor unsupported by lack of adequate standards.

Leverage at the interfaces introduces the notion that the actual architecture of a collaborative SoS resides at the interface specifications. When designing a system for interfacing within a SoS, Rechtin and Maier suggested the design team to choose standards that maximize the opportunities for the participants to find individually beneficial strategies.

The *Ensuing cooperation* theme is closely related to the existence of stable intermediate forms. Chattopadhyay's perceived net benefit discussions are the formalization of this principle; which means individual systems need to perceive value for themselves in the cooperation to engage.

This section reviewed relevant techniques for architecting SoS, which will be summarized again at the end of this chapter. Before, we need to introduce a few additional methods that will prove useful to address the research questions of this thesis.

4.4 Additional methods literature

Besides systems architecting techniques, three specific methods which are related to the thesis' approach are presented: Markov decision chains and Time-expanded Decision Networks (TDN). Moreover, we briefly discuss Network flow maximization problems as they a useful modelling approach to allocate resources in a FoS.

4.4.1 Network flow maximization problems

Maximum network flow problems are regularly applied to a variety of problems including bandwidth allocation, electric grids, water supply systems and space logistics (Takuto Ishimatsu et al., 2013). In its classic form, the problem addresses the allocation of flows between a set of nodes with different edge capacities, so that the flow between a source and a sink is maximized.

As any SoS, FoS are a net-centric entity with dynamic resource allocation and exchange between systems -or nodes in the network representation. As such, network flow problems are common and need to be understood and solved. The first and most famous algorithm to solve flow maximization problems is the Ford and Fulkerson algorithm (1956). In its basics, the Ford & Fulkerson algorithm is based upon progressively augmenting the flow units through a path (set of edges) from the sink to the source, until no more augmenting paths exist.

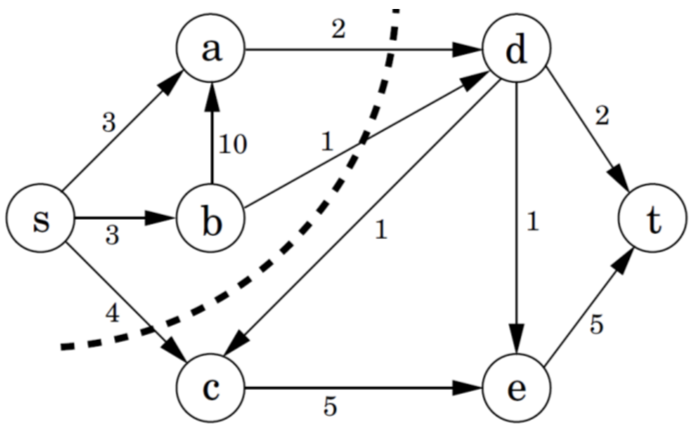

Figure 10. A network represented by a graph with source s, a sink t, nodes S={a,b,c,d,e} and edges E={{s,b},{s,a},{s,c},{a,d},{b,d},{c,d},{d,e},{d,t},{e,t}} from (Vazirani et al., 2006). Each edge displays its flow capacity (e.g., litres of water in a pipe system). The cut shown in dashed line with S={a,b} is the minimum cut, hence the maximum flow reachable is 4+1+2=7 units.

Figure 10 illustrates a network representation as a graph. The maximum flow found by Ford & Fulkerson or other algorithms is proven equal to the *minimum cut* capacity of a given network. Such correspondence is known as *the max flow-min cut theorem*, (Bertsimas and Tsitsiklis, 1997) which is the basis for optimality guarantees.

A cut is a subset of nodes in the networks under consideration, which has a bounded flow capacity to the rest of the network. This capacity is the aggregate of the capacities of the arcs that have been disconnected, or cut from the remainder of the network as shown in Figure 10. The cut capacity is hence a bottleneck, and finding the most constraining of this bottlenecks –the minimum cut– is equivalent to finding what is the maximum flow traversing the network from the source to the sink.

Despite the existence of specific algorithms for maximum flow problems, many practitioners employ Linear Programming (LP) formulations to solve flow network problems (Vazirani et al., 2006). LP is a general method which enjoys widespread adoption, a rich literature and many supporting solvers. In LP theory, the max flow-min cut theorem is understood as an instance of the LP duality theorem (Vanderbei, 2013). Typically, when formulating a maximum flow problem as an LP, we take as independent variables the flow allocations on each edge, and we attempt to maximize a target function which adds up the flows incoming to the sink, or leaving the source. We need to define the equality constraints so that they enforce the conservation of the incoming and outgoing flows at each node. The role of the inequality constraints is to prevent the flows from exceeding the total capacity of each network edge.

LP for maximum flow problems has already been applied by the author to FSS, as part of a first-pass framework preparatory to this work (Lluch and Golkar, 2015). For the exercise proposed in (Lluch and Golkar, 2015), the FoS was interpreted as a multi-commodity network (Takuto Ishimatsu et al., 2013). Commodities represent the data generated in each federated system, which was deemed of different value and hence needed to be treated independently. The goal was then to set the different commodity flows for maximum retrieved value. The framework proposed herewith will leverage on such techniques to optimally allocate resources in an FoS and understand potential benefits of federating.

4.4.2 Markov Decision Processes

The goals we have set regarding the evaluation of federation evolution and emergence require of the inclusion of techniques able to capture a time aspect or sequences of events. *Markov Processes* (MPs) can support such analysis. A Markov process (Feinberg and Shwartz, 2012) is a sequence of random states with the Markov property, i.e., the probabilities of transitioning to a future state only depend on the current state of affairs; the process is memoryless. Many problems can suit this definition, if formulated adequately. When the set of events under consideration are discrete, the Markov process receives the name of Discrete-Time Markov Chain (DTMC), as opposed to a continuous time Markov process. Moreover, if at each state we can take decisions about where to progress next, that affect the probability distribution of states on the next step, we are talking about Markov Decision Processes (MDP). Many tree search problems can be formalized as an MDP. What configures a Markov Decision process is a tuple $\{S,A,P,R,\gamma\}$ (Silver, 2015) where:

- S is a finite set of states representing a process,
- A is a finite set of actions that can be taken at each state,
- P is a state transition probability matrix, composed by:

$$P_{ss'}^{a} = \mathbb{P}[S_{t+1} = s' | S_t = s, A_t = a] \qquad \text{Eq. 4-13}$$

Which reads as the probability for a future state in time $t+1$ to be s', given that in current time t the state is s and we take action a.

- R is a reward function:

$$R_s^a = \mathbb{E}[R_{t+1}|S_t = s, A_t = a] \quad \text{Eq. 4-14}$$

Which reads as the expectation of a reward in state $t+1$ given the current state and an action a,

- γ is a discount factor between 0 and 1, which discounts the yield of future rewards.

MDPs are usually represented by a graph where the states are nodes and the edges represent transitions between states with their assigned probabilities. Additionally we can attach the decisions to each state. Figure 11 illustrates an MDP with 4 states, and two decisions associated to state *S1*.

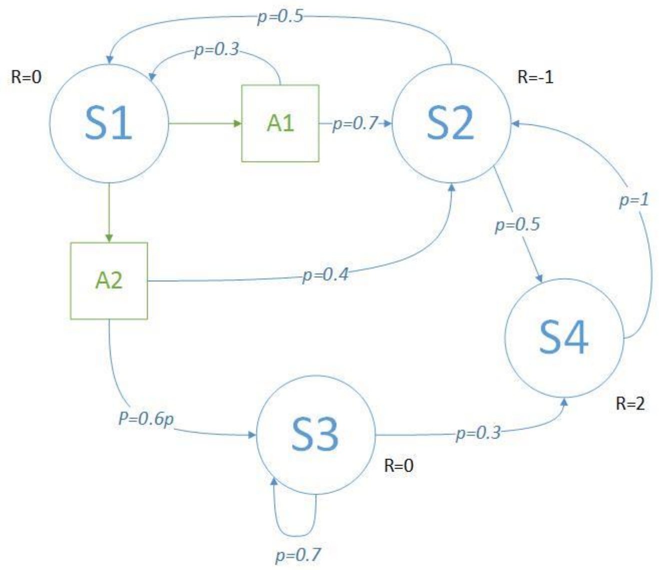

Figure 11. Example representation of a MDP with 4 states, 2 decisions attached to S1, a reward for each state and state transition probabilities.

MDPs are widely used to model the world and decisions in agent planning (Mausam and Kolobov, 2012) problems, where the goal is to find an optimal set of actions for the agent to pursue such that the expected outcome attains the highest rewards. We call an *episode* a specific sequence of states. For instance, S1→S2→S4→S2→S1→S2 is an episode of 5 timesteps. In this episode, and with an aggregated rewards model, the agent navigating this MDP would have attained 0+(-1)+2+(-1)+0+(-1)=-1 reward. Formally, the rewards are attached to all the edges –the reward is attained when transitioning from s to s'– but for simplicity, in Figure 11, we assigned rewards to states, achieved when the given state is reached from any edge.

MDPs might have or not have not any terminal states or exit nodes, potentially going on forever as in the example above. If the episodes can be infinite, it is no longer possible to compare them in terms of achieved rewards, since the latter can be infinite. An approach to save this caveat is to apply a discount factor γ, to switch preference to earlier rewards. Such discount is to be applied per episodic timestep, as in Eq. 4-15.

$$R^a_{total} = R_1 + \gamma R_2 + \gamma^2 R_3 + \gamma^3 R_4 + \cdots = \sum_{t=1}^{\infty} \gamma^{t-1} R_t \leq R_{max}/(1-\gamma) \qquad \text{Eq. 4-15}$$

Computing this infinite geometric series for $\gamma<1$ we obtain a bounded reward as the rewards in long-term future have very little influence. Alternatively, we can evaluate bounded futures (i.e, up to 5 states in the future) as another way to deal with MDPs. Provided we can evaluate rewards in future scenarios, the next question to ask ourselves, and the core problem of MDPs, is what is the expected future rewards when the system or agent is in a given state, and what actions can we take to maximize such rewards.

In an MDP, A *policy π* is a mapping of States to Actions, S→A. Each action has an expected utility, called $q^*(s,a)$, which is computed based on future rewards, assuming all posterior actions are optimal. Hence, we want, at each decision step (the square nodes in Figure 11) to pick the action associated to the maximum $q^*(s,a)$. This implies being able to compute the future rewards achieved as discussed above. We associate a value V^* to each state based on the $q^*(s,a)$:

$$V^*(s) = \max[q^*(s,a)] \qquad \text{Eq. 4-16}$$

$$q^*(s,a) = \sum_{s'} P(s,a,s')[R(s,a,s' + \gamma V^*(s')] \qquad \text{Eq. 4-17}$$

Eq. 4-17 illustrates how to compute the $q^*(s,a)$ based on the value of posterior states. Eq. 4-17 states that the maximal expected value of taking an action a in a state s corresponds to the rewards of achieving state s' and the discounted posterior rewards $V^*(s')$ of following optimal decisions. This is summed over all the potential successor states s', weighted by the s to s' transition probability $P(s,a,s')$. Substituting Eq. 4-17 in Eq. 4-16 we obtain a compact, recursive expression, which is a form of the Bellman equation (Silver, 2015):

$$V^*(s) = max \sum_{s'} P(s,a,s')[R(s,a,s' + \gamma V^*(s')] \qquad \text{Eq. 4-18}$$

Therefore it is possible to compute the values of each state based on current and future probabilistic rewards, and conduct policy searches that yield the best-possible decision making. The most common method to solve the Bellman equations for MDPs is the value iteration method which implies starting in a set of state in a particular timestep k, assume a V_k^* value for them and iterate backwards, computing V_{k-n}^* value based on previous states, up to the present state. Expanding the k timestep to further future states generally leads to convergence, thanks to γ discount factor. The time complexity of this procedure is $O(nS^2)$, S being the states and n the iterations considered.

Obviously, when the policy search space is large (many actions possible) and the amount of states is also large, the computation time required can be long or even unfeasible. However, even if not computing full optimal policies, MDP formulations allow evaluating reduced problems of comparing specific policies in a finite time-horizon, which is also informative for agents or systems navigating an MDP.

Is not always possible to fully characterize the rewards, the underlying states or the transition probabilities in an MDP. Approaches to deal with imperfect information in MDPs include Partially Observable MDPs (Monahan, 1982) and fuzzy MDPs (Kurano et al., 2003).

4.4.3 Time-expanded Decision Networks

Time expanded Decision Networks (TDN) is a methodology developed by Silver and De Weck (2007) to support strategic decision-making for evolvable, complex systems suitable of multiple configurations. TDN builds on, and is also an alternative, to real options analysis (Saleh et al., 2002) as means to evaluate the value and of transitioning to alternative system configurations and uncertain futures, and quantify the costs associated to embed such potential transitions in the design.

The starting point of TDN is a *static network*, represented by a graph where nodes are specific system configurations, and arcs represent feasible transitions between system configurations. Each transition arc has a cost assigned to it which represents the switching costs of transitioning from one system configuration to another.

Moreover, the operating costs of continuing to operate a specific configuration are also included as an arc with its origin and destination in the same node. Finally, the design of the system and the notion of decommissioning are represented by a source and sink node, respectively. The former has also a cost model accounting for design costs of each system configuration.

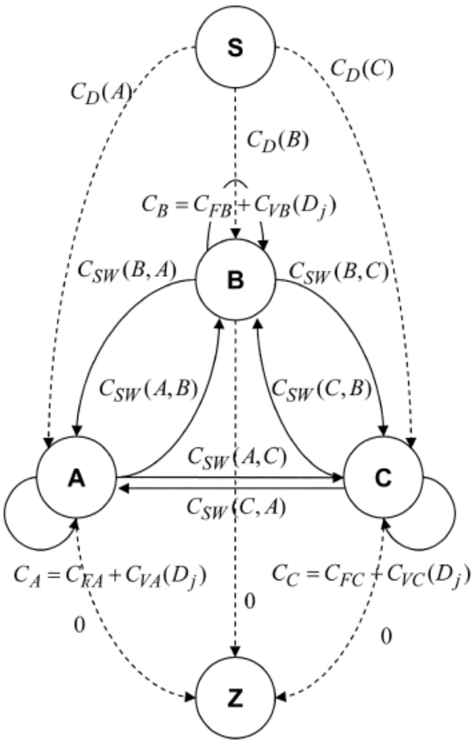

Figure 12. A static network depicting 3 system configurations A,B,C, a Source S and a Sink Z, from (Silver and De Weck, 2007). C_{sw} represent switching costs between configurations and C_D are design costs. Each node has an arc to itself representing the continued operation costs, from which C_F are fixed and C_v are variable.

Figure 12 represents a static network with 3 configurations A,B and C. Recognizing that the feasibility of transitions and its associated costs change with time, Silver and De Weck build a time-expanded network on the static one based upon discrete time states. On such a network is then possible to apply any shortest path or minimal flow solvers (Ford Jr and Fulkerson, 1956) to find the paths that minimize lifecycle costs by adequately cycling –or not– between system configurations.

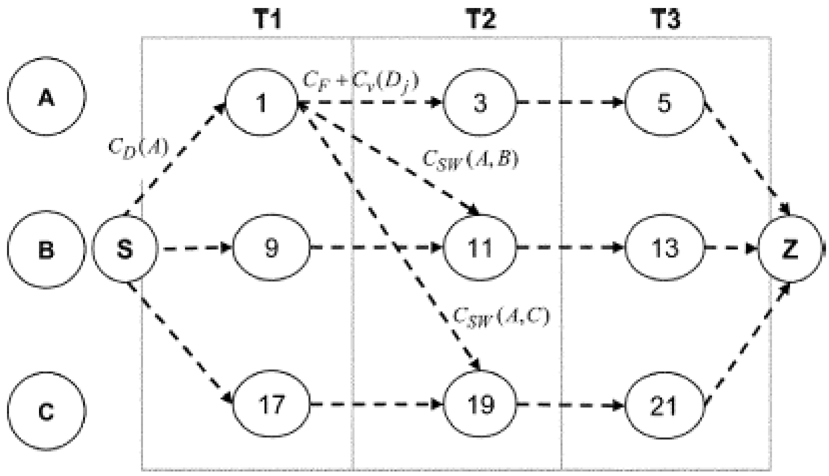

Figure 13. A TDN based on the previous example, with 3 time steps. Only at time step t1 it is possible to switch from alternative A to alternatives B and C (arcs 1-> 11 and 1->19) from (Silver and De Weck, 2007).

Figure 13 illustrates a TDN based on the static network discussed previously. At each time step T, which is conceptually very similar to the notion of epochs defined by Ross (see 4.2.4) the transition costs and the demand scenarios might change. That is, the level of service that is required from the system changes, making some configurations perhaps more desirable than others in the measure that they can serve this demand increase with less cost per unit. This is the mechanism used to stimulate changes between alternatives under non-zero switching costs, and is also the way Silver and De Wecks' formulation captures the notion of benefit of alternative configurations; as a response to exogenous demand changes. This is resonant with the classic definitions of flexibility as the ability of a system to adapt to demand changes. Such definitions and metrics can be found in (Nilchiani, 2005). TDNs have been applied to a heavy lift launchers planning example. While not shown in Figure 13 for simplicity, TDNs formally include decision nodes which allows to decouple switching costs arcs from continuing-operations costs arcs.

The TDN method belongs to the decision-making support field and therefore has commonalities with decision tree analysis (Nilchiani, 2005) and MDPs. While the amount of timesteps, demand scenarios and alternative configurations considered could be virtually infinite, a deliberate reduction of those renders a TDN the minimum cost path computable in reasonable time. The time complexity of the method when solving for shortest paths depends on the configurations, time steps and different demand evolution scenarios as $O(SC^2T)$. If we interpret the amount of timesteps as the amount of states downstream considered by an MDP, and the configurations as MDP states, we will see that the complexity is identical to MDP's $O(nS^2)$ when solving for an optimal policy.

TDN has been tailored to compare, on cost, series of system evolution paths under different demand scenarios. However it does not include explicit measures of increased utility or benefit of adopting system alternatives which can be of interest when architecting systems, rather choosing to embed the benefit as a cost-effective reaction to demand increase. This approach is advantageous since it can consolidate costs and benefits in a single unit.

4.5 Distributed Space Systems and Federations

This section reviews both novel DSS, including federations and classic space distributed systems, and provides a historical perspective on their emergence.

Constellations are the first type of distributed space systems conceived and deployed. Initial research on constellations dates from the 60s, under the generic title of satellite networks. Luders (1961) started developing algorithms in the 60's to achieve desirable earth coverage patterns with a group of satellites. This research was progressively generalized and improved during the 70's and 80's, with the works of Walker (1977), Ballard (1980) and many others (Beste, 1978). At the same time, the first constellations were being deployed (Martel, 1981). An excellent literature review on satellite constellations alone can be found in (Ha, 1998). Satellite constellations are a mature space architecture that emerged as a solution for a well-defined spatial and temporal coverage problem, which had an intuitive solution on distribution. Availability and reliability are topics essential to constellations; availability is usually formulated as explicit system requirement, while reliability must provide for it, through redundancy –spare satellites, either on the ground or on parking orbits–.

While the development of satellite constellations responded to primary functional requirements, such as continuous worldwide coverage, the novel DSS are better defined as an approach rather than an answer to explicit requirements. The advent and success of the satellite constellations paved the way for the apparition of more advanced distribution concepts, which progressively moved from requirements-oriented to value driven approaches. The loose term 'satellite clusters' started appearing in the 80's (Renner and Nauck, 1984) to address groups of spacecraft flying arbitrarily close, working together for the same goal. Geosynchronous broadcast missions were the initial application niche envisioned for clustering. The only clear boundary between clusters and constellations that would stem from the papers on satellite clusters is the existence of a permanent line of sight between elements in clusters. That is, all spacecraft in the cluster shall be able to see any other at any time, ruling out central body obstructions of inter-satellite links which are common in constellations.

Satellite clusters are of particular interest for Synthetic Aperture Radar missions (SAR). Multiple research papers examine satellite clusters for interferometry SAR applications (Stiles et al., 2000). The tight requirements on spacecraft relative positions for SAR and similar missions led to the use of the 'satellite formations' (Bristow et al., 2000) term, to reflect the need to keep precise guidance and navigation. This is only relevant as opposed to satellite swarms (Engelen et al., 2011), a more recent version of a cluster where spacecraft orbit in loose formation with less stringent position requirements.

After the appearance of formations and swarms, the cluster designation was kept as a general, catch-all term. A special case of swarms are Space-Based Wireless Sensor Networks (SB-WSN) (Vladimirova et al., 2008), which are the application of WSNs to space, based upon very small sensor nodes to replace a larger spacecraft.

Formations and Swarms have a bottom-up, technology-driven literature, closely related to the small satellite community. Partially resulting from the lack of ultimate applications and clear top-down mission needs, satellite formations and swarms researchers advocate for them via system properties such as flexibility, agility and re-configurability, generally called *ilities* (de Weck et al., 2012). These are characteristic advantages of formations and swarms when compared to monolithic systems. However, they are difficult to assess quantitatively. This turn to value-driven architectures was completed with the rise of Fractionated Spacecraft concept (Brown and Eremenko, 2006).

Fractionated Spacecraft propose to break down a conventional spacecraft into physically detached subsystems and components, which fly in close proximity. Fractionated spacecraft are a new architectural solution to the old problems posed by monolithic design, such as requirement coupling and design rigidity. While Swarms and Formations are generally composed of identical or very similar spacecraft, a Fractionated Spacecraft is heterogeneous per se as different spacecraft functions are allocated to different free-flying fractions. This presents component inter-dependency which is not necessarily the case in Formations and Swarms. Total component independency is only found in the latest of DSS proposals, FSS. Figure 14 introduces the timeline of the different DSS introduced here.

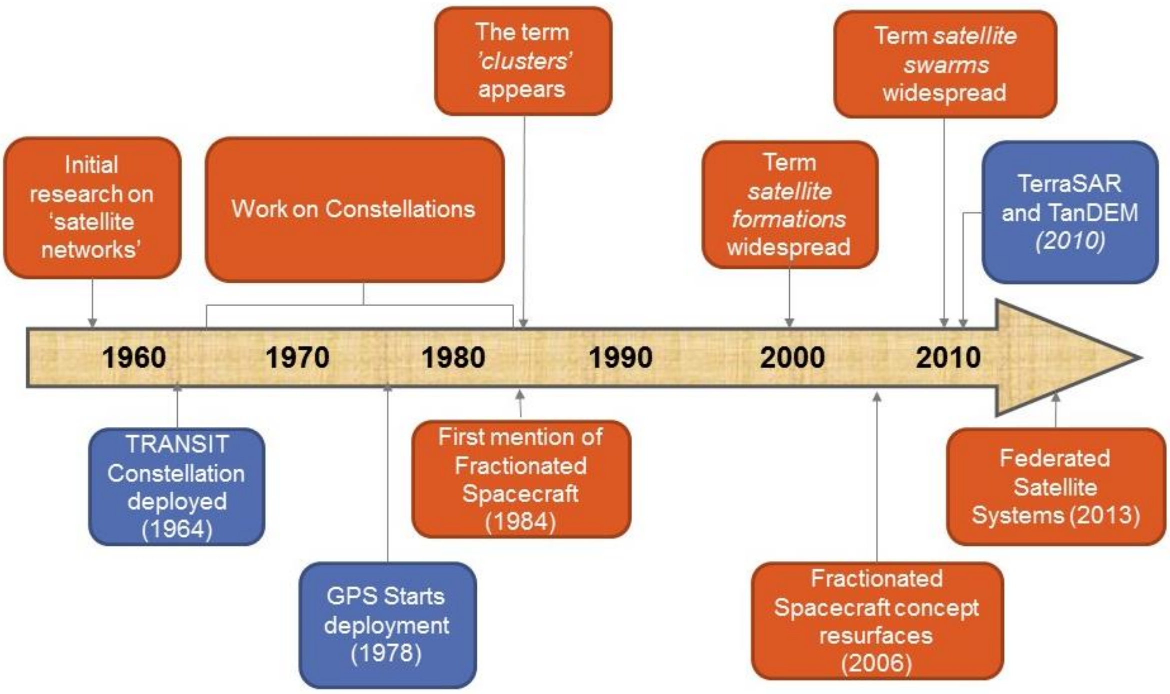

Figure 14. Timeline for DSS literature, shown in context with the deployment of TRANSIT, GPS constellation and TerraSAR-TanDEM formation flight ("Eoportal missions directory," 2016).

FSS have been introduced in chapter 2 and are the main motivation of this work. FSS aims to transition from the paradigm of independent, isolated operation of spacecraft to a cooperation environment where spacecraft are still independent and respond to different goals, but can exchange resources as needed. In line with the concept of federations, this includes data-based resources such as downlink bandwidth, storage and processing. This requires the realization of opportunistic, multi-stakeholder space networks.

A similar concept to FSS was also theorized by NASA Ames and JAXA researchers in a short series of 2014 papers under the name of Heterogeneous Spacecraft Networks (HSN) (Faber et al., 2014). HSN research was very short lived but it provides a form of idea validation and is informative of its interest of making a more effective usage of assets by agency stakeholders. Both FSS and HSNs propose an interoperable, networked approach to space missions. FSS literature differs from HSN and general DSS literature in its focus. While technology aspects are not neglected, FSS work focuses deeper on illustrating applications, case studies and design tradeoffs. Table 8 summarizes the characteristics of the DSS introduced here.

Table 8. Distributed Space Systems comparison. Mission references can be found in ("Eoportal missions directory," 2016).

Architecture concept	Mission goal	Element uniformity	Element autonomy	Geometry configuration	Concept maturity	Year Proposed
Satellite Constellation	All elements contribute to the mission objective	Homogeneous, while different variants are possible (GPS generations)	Independent elements	Structured ~ thousands of kilometers ranges	Many deployed constellations: GPS, Iridium	60-70s'
Satellite Formation	All elements contribute to mission objective, individual objectives possible	Homogeneous elements – functional differences possible	Autonomous	Tightly-controlled close flight – from meters to kilometers range	TandemX and TerraSaR, FASTRAC	Late 90's
Satellite Swarm	All elements contribute to mission objective, individual objectives possible	Homogeneous elements – functional differences possible	Autonomous	Loose, close flight - from meters to kilometers	Demonstrator missions: ESA´s SWARM	00's
Fractionated Spacecraft	All elements contribute to mission objective, individual objectives possible	Heterogeneous elements	Inter-dependent elements	Close flight- from meters to kilometers	Pending demonstrator mission	1984 & 2006
Satellite Federation	Unrelated mission objectives, ad-hoc cooperation	Heterogeneous elements	Independent elements	Unstructured thousands of kilometers ranges	Pending demonstrator mission	2013

By the time of the conclusion of this thesis, the author is aware of only two works addressing the classification of DSS including the most recent concepts proposed (Poghosyan et al., 2016; Selva et al., 2017). Such publications are very recent and this thesis' author is involved in both as co-author. This section provides solely the authors' perspective on the subject notwithstanding similarities and differences with the aforementioned papers.

4.5.1 Overview of communications network technologies

For satellite federations, and also in other cases, a communications network is established amongst the participating systems. We very briefly touch upon relevant technologies here, including network and physical layer aspects. The interested reader is referred to (Lluch et al., 2015) for a detailed example of a communications network implementation in FSS. We extract here some key points from that work.

The voluntary nature of a system's participation in a FoS can, in the worst case for communications, render unpredictable and/or unreliable network topologies. This could require of network implementations supporting autonomous, node-to-node network discovery and yield a dynamic network topology. It is of interest then to explore some fundamentals of Mobile Ad-hoc Networks (MANETs) (Hoebeke et al., 2004).

4.5.1.1 Mobile Ad-Hoc Networks (MANETs)

MANETs are meshed networks where the component nodes do not make use of an existing infrastructure, instead communicating with the other nodes in range directly via wireless links. In order to reach nodes outside the communication range, intermediate nodes perform routing functions and relay packets for others. Thus, the nodes in a MANET handle themselves the routing and discovery of the network.

Routing protocols in MANET operate with incomplete information about the network topology and node status. Routing in MANETs is an iterative process where nodes only attempt to send the packet in the right direction on first hop, instead of prescribing a full route which could be later be unavailable. The most popular MANET routing protocols include Optimized Link State Routing (OLSR) (Huang et al., 2008) and Better Approach to Mobile Ad-hoc Networking (BATMAN) (Delosieres and Nadjm-Tehrani, 2012). Both are used in the WCNs described in Chapter 8. Nodes both in OSLR and BATMAN send information of their status to 1 or 2-hop neighbors, maintaining only partial network information instead of a full routing table. Huang et. al. (2008) investigated the route stability and duration of an implementation of OSLR in MANETs, and pointed minimum-hop counting as an insufficient metric for path computation. Their work exemplifies how in unpredictable, rapidly changing MANETs, link availability is the main concern and driver of performance. The mechanism to take this into account is to use link availability, or connectivity weighting (Jiang et al., 2005) when computing shortest paths. Delosières and Nadjm-Tehrani (2012) integrated the store-and-forward concept to BATMAN, paving the way for the use of the BATMAN features in disruption-prone environments such as space, hence being applicable to FSS.

We can draw solutions for FSS and other FoS communications at the network layer from MANETs. However, we must also consider the limitations of systems at the physical layer. This concerns the system's communication equipment interoperability (Don and Sage, 2012). The literature on FSS considers the usage of a standardized ISL as the simplest technology solution for federations. However, it has also reflected on the usage of flexible communication equipment capable of adapting to different physical layer standards; with research and experiments on Software-Defined Radios (SDR) (Akhtyamov et al., 2016).

4.5.1.2 Software Defined Radios

A SDR (Angeletti et al., 2014) is based upon a Digital Signal Processing (DSP) component and RF front end. An analog-to-digital converter allows for most operations to be performed on the digital domain. The DSP is typically performed on a programmable FPGA, the key of the radio's flexibility. Some hobbyist applications are possible by just adding an RF front-end to a PC sound card, including receiving satellite downlinks (Maheshwarappa et al., 2015). The flexibility of an SDR is limited by the radio front end, especially on the antenna side, where only a limited range of frequencies, usually within a band, is available. Workarounds to this can include more complex systems with multiple or reconfigurable antennas.

The first SDRs appeared on the 90's for military use (Mitola, 1992), and have progressively made their apparance in the commercial market. SDRs can adjust their carrier frequencies and change their communications protocols on the fly via software changes. A system equipped with an adequate SDR could, to a certain extent, interact with several existing systems and even allow them to interoperate by acting as a network mediator. This could theoretically allow a federation to include systems already deployed for which changes are not possible, or upgrade its network infrastructure via software. This technological solutions can enhance the FoS evolution management and expand its adoption base.

4.5.1.3 Physical Layer: Inter-Satellite Links

ISL are the media of the proposed data exchanges between FSS participants, the physical layer of the FSS network. Since the initial experiments in the 70s, ISLs have been deployed in several constellations and other DSS types referred in this section.

ISL in medium and large satellites has been a mature technology for more than 30 years. For nanosatellites (<10 kg) ISL is still indeed maturing and only the last years have seen an increase in technology demonstrators and projects attempting to fit ISL capabilities in the constrained volume and mass of nanosatellites (Muri and McNair, 2012). The most mature technology for ISL is RF. Typical RF performances support datarates of 50Mbit/s and more at 5,000 km link distance (Selva et al., 2017). For nanosatellites, where the antenna size and transmit power are limited, the achievable performances are typically hundreds of Kbit/s at 5,000 km (Selva et al., 2017). Besides the corresponding power and mass allocation requirements in the spacecraft, including ISL capabilities can have impacts in the attitude control, on-board data handling, thermal subsystem and spacecraft decks configuration (Lluch i Cruz and

Golkar, 2014). Table 9 presents a brief selection of satellite systems that have used ISL, including the mission type, the ISL technology and the deployment year.

Table 9. Selection of missions with ISL capabilities for demonstration or operational purposes. Data from ("Eoportal missions directory," 2016) and (Muri and McNair, 2012).

Mission	Mission type	ISL technology	Year commissioned
OSCAR 6-8	Technology demonstrator	VHF band	1972-1978
TDRSS	Relay system	C, K, Ka band	1983
Iridium	Communications	K band	1997
Navstar block IIR	Navigation	UHF band	1997
SPOT-4	Earth Observation	Optical	1998
Artemis	Technology demonstrator	Ka, S band, and optical	2001
SAR-Lupe 1-5	Earth Observation	S band	2006-2008
CANX-4, CANX-5	Technology demonstrator (nanosatellites)	S band	2014
Sentinel-1,2	Earth Observation	Optical	2014, 2015
EDRS	Relay system	Ka band, and optical	2016

Optical communications (Hemmati, 2006) are also used in ISLs. Albeit less mature than their RF counterpart, optical communications show great promise and are experiencing intense development efforts nowadays (Edwards, 2014). ISLs based on optical communications can boast datarates in the order of Gbit/s, with reduced power and mass footprint respects to RF. However, optical ISLs require a pointing accuracy in the order of arcseconds (Selva et al., 2017). Such performance is attainable for conventional spacecraft but it is yet to be seen if nanosatellites can achieve it in operational conditions (Kingsbury et al., 2015). ISL is typically used to relay mission data using dedicated services, to provide ground communications routing, to receive telemetry and upload commands to the spacecraft, and for specific functions like ranging and ephemeris propagation as in the case of Navstar (M. Abusali et al., 1998). We will see more cases of ISL usage in the future, as several of planned constellations (Selva et al., 2017) include ISL to provide communications relay services to ground users, or to grant constant access to the spacecraft.

4.6 Literature survey summary and research gap

This chapter has reviewed the SoS literature, established FoS in its context, and discussed the principles of federalism applicable to engineering systems. Moreover, it has introduced the systems architecting discipline, surveyed the most relevant methods for SoS architecting, and the underlying theory. Finally, we have covered specific methodologies to support the research approach.

SoS have existed as an academic field for approximately 20 years. During these two decades, most of the literature has been concerned with appropriately identifying the characteristics and differentiating elements of SoS. A number of frameworks for architecting SOS exist. Two notable ones quantitative ones are Ricci's SAI method and Chattopadhyay's SoSTEM method, discussed in this chapter. These two frameworks, and all of the rest, reason on a global tradespace representation, explicitly assume that there is a global SoS stakeholder with influence on the SoS acquisition and deployment process, and the SoS, while evolvable, is deployed at once, or on a coordinated schedule. These assumptions match very well directed SoS but are not adequate to the object of this thesis; these works attempt to support decision-making for acquisition, while for FoS, the key question to answer is if, system by system, it makes sense to architect for federation.

The independency of the federated systems and the feedback loops between their value delivery to stakeholders, require a different perspective for quantitative analyses. As discussed in the SoS definitions section, FoS fall under Maier's definition of virtual SoS, and to analyze those we need to focus on the non-hierarchical relations between systems in the absence of centralized goals and management structures. Notwithstanding the focus and assumptions differences, the surveyed works are fundamental at the methodological level. A first pass framework for FoS similar to said SoS frameworks (Lluch and Golkar, 2015) was developed by the author of this thesis as an exploratory step leading to the current work. In that previous work, we studied different architecting alternatives for an Earth Observation missions' federation, that is, a federation of satellite systems. The analysis implied the combination of utilities of different missions into a global utility-cost tradespace to quantify the benefit of federating. As discussed before, we recognize the limitations of such an approach in FoS where 1) global tradespace representation can be a means to an end but is not conceptually legitimate due to the absence of global decision-making stakeholders, 2) aggregating utilities from different missions is troublesome as it implies a prioritization or preference between missions that cover different EO aspects for different stakeholders. These caveats, that are not critical to directed SoS, are critical to virtual SoS/FoS and to be addressed in this research. Table 10 details the differences between what existing frameworks target and what is needed for FoS.

Table 10. Detailed comparison between existing approaches and what is needed for FoS.

<table>
<tr><th>Method</th><th>Design Object</th><th>Focus</th><th>Supports decision of</th><th>Global mission objectives</th><th>Analysis of SoS evolution</th></tr>
<tr><td>SAI [Ricci et al., 2014]</td><td rowspan="4">the SoS</td><td>Analyze and enhance SoS properties ('ilities')</td><td rowspan="4">SoS architects, and SoS stakeholders</td><td rowspan="4">Yes</td><td>Using MATE-RSC and perturbations</td></tr>
<tr><td>Robust optimization framework [Davendralingam and DeLaurentis, 2013]</td><td>Network architecture optimization for robustness</td><td>No</td></tr>
<tr><td>Integrated SoS Synthesis [Sobieszczanski-Sobieski, 2008]</td><td>SoS hierarchical decomposition and optimization</td><td>No</td></tr>
<tr><td>SoStem [Chattopadhyay, 2009]</td><td>Comparing the utility of several SoS architecture alternatives</td><td>Using Epoch-Era analysis [Ross et al., 2008]</td></tr>
<tr><td>A method for FoS?</td><td>Each individual system</td><td>Analyzing the benefits of federating for each system</td><td>Architects and stakeholders of individual systems</td><td>No global tradespaces or mission statement</td><td>Needs to capture the specifics of deployment and value change through time</td></tr>
</table>

Besides the specific SoS architecting methods, we reviewed architecting methodologies of general application like MATE –on which some SoS frameworks are based upon– and TDN. Both bring into systems architecting the dimension of time and an evolving operational context. This dimension is also of utmost importance for any work on FoS, as a federation is by definition an evolving subject. TDN and MATE effectively analyze alternative embodiments of a single system, which has a defined goal and stakeholder set. While the multiple applications of MATE have dealt with endogenous system changes and also SoS, unfortunately they do not support analysing the specific architecting challenges for FoS. For FoS, we need to explicitly analyze the coupling effects on architecting decisions on more than one system, where these systems are disparate in goals and have in general non-aggregable utilities.

Hence, the architecting process is no longer subject to a single viewpoint, but there are as many architecting problems as systems in the federation, each own with their own utility-cost functions. Table 11 summarizes this discussion and connects it to the relevant literature themes.

Table 11. Summary table of literature themes, and research subject differences.

	Conceptual aspects	**Approach to formation and adoption**	**Long-term evolution aspects**
Existing SoS work	Global stakeholders, hierarchical structure	Assessment of global value	Time change of stakeholder's utility function(s) and operational contexts
FoS	Multiple stakeholders, operational and managerial independence	Focus on identifying synergetic configurations	Assess the value for the inidividual systems as other systems join or abandon the federation
Related literature themes	SoS typologies, virtual SoS, Principles of federalism	Utility theory, MAUT, MOO, tradespace exploration, Multi-commodity Network flow optimization	Dynamic MATE, TDN, Markov Decision Processes

Leveraging on the techniques surveyed here and aware of the specific requirements of architecting FoS, the next section describes the proposed approach to answer the research questions in this thesis.

Chapter 5 Approach

This chapter introduces the approach to answer the research questions, structured as an architecting framework. First, we need to establish a series of common definitions and lexicon to to structure our disquisition about federations. Then, this chapter introduces in more detail the concept of synergy, reasoning through the concepts of utility and tradespace exploration. Next, we describe the steps of the architecting framework, which is based on a *static* and a *dynamic* aspect to accomplish its purpose. To conclude, this section discusses the role of incentives, coordination mechanisms and authority in FoS.

5.1 Fundamentals: defining federations

The definitions of Maier for virtual systems and Sage and Cuppan for federations, discussed in the literature review chapter, are the essential cornerstones we use for this thesis' definition of federation. This thesis considers a ***federation*** to be:

A set of systems with independent goals, management and operations, that possess the adequate interfaces to cooperate and do so when it is advantageous for all parties involved.

Besides this definition, we adhere to three of Sage's principles: *subsidiarity, a uniformed way to do business, and dual citizenship*. The principles of *Interdependence* an*d Separation of powers* however do not add much to this definition for engineering systems. The basic notion conveyed here is that the systems within a federation can carry on their purpose without a need for the other systems; however, federating might pose additional advantages in fulfilment of each systems' purposes. While built for engineering systems, the definition above is also fit to describe sharing economy environments.

Each system in the federation has a ***local***, or individual perspective, driven by the views of its stakeholders and beneficiaries (Crawley et al., 2016a) and the value the system delivers to them. As per the meaning of *system*, we observe through this text the definitions proposed in chapter 1. Each system in the federation is assumed to have its architecture, which in general is different from any other system in the federation. Any architecting tradespaces of the system, when examined before any consideration to federation, we will call ***standalone tradespace***s, and the corresponding Pareto front(s), when applicable, we designate as the ***Standalone Pareto Front(s)*** (SPF). In this chapter we provide examples of both. In this work, we name ***FoS configuration*** the specific arrangement of the FoS, including the particular architectures of all the participant systems together with the interfaces in place. A FoS configuration captures all the necessary elements to distinguish a particular FoS embodiment from another. We will call the collection of different FoS configurations in time a ***FoS configuration path***.

We call the systems in the FoS simply ***participant systems***, or ***peers***. The operations that realize a FoS are the accesses to other peers' capabilities, denominated ***resources*** here. We consequently recognize the *raison d' être* of an FoS is to re-allocate resources amongst systems such that all those involved are benefited. The resources are owned by individual systems, and are exchanged on a commercial (via incentives) or a voluntary basis. The nature of the resources can vary, including accesses to an external infrastructure, to information, or to specific capability.

5.1.1 The processes and operands of cooperation

The proposed definition of FoS includes the concept of cooperation between engineering systems, during their respective operations. Now, let us briefly delve into the technological forms, and the functions performed through cooperation in nowadays technological environment. As a systemic process, cooperation requires an operand (Crawley et al., 2016b). Crawley classified operands in *information, matter,* and *energy*, and included data, commands, and thoughts as subclasses of the information operand.

In the current technological context, it is straightforward to envision *information and data* exchange between systems. Digital communication technologies are pervasive and have enabled paradigms like the IoT and M2M communications. Systems that fall into FoS definition and perform data-based exchanges include Vehicle to Vehicle (V2V) networks (Gerla et al., 2014), Wireless Community Networks (WCN) (Vega et al., 2012) and FSS. The last two items are analyzed as case studies in this thesis, while we also briefly discus V2V. Based upon data operands, the resources or capabilities that can be accessed cooperatively are bandwidth, storage, computing power, and the data itself for enhanced operational awareness.

We can illustrate the use of *energy* as a cooperation operand in the case of Vehicle-to-Grid (V2G) applications, which are part of the smartgrid paradigm (see annex 1).

Finally, *matter* exchanges are more challenging to envision as the key cooperation operand in FoS. Interfaces for matter exchange (such as flow canalization, freight transport) are better understood as an infrastructure or system of their own than an interface. Even if considered so, they are interfaces between social and economic actors, rather than between engineering systems. Moreover, mutually advantageous mass exchanges respecting managerial and operational independence are difficult to envision. We might be able to identify matter exchanges between systems in very specific cases of directed SoS, such as space logistics networks (Takuto

Ishimatsu et al., 2013). Table 12 summarizes the cooperation processes, associated benefits and traits of FoS.

Table 12. Operands for cooperation, associated processes and interfaces for FoS.

Operand	Cooperation Processes	Interfaces	Example FoS
Data	Relay for others	Communications Network and protocols	FSS, V2V, WCN
	Store for others		FSS, V2V
	Process for others		FSS, V2V
Data Content (Information)	Exchange for operational awareness		V2V
Energy	Exchange, Balance grid load	Power cables, bi-directional chargers	V2G
Matter	Transport	Flow canalization, freight transport	?

In this thesis' perspective, the objective of cooperation is to support the local goals of the systems involved, instead of an overarching SoS goal. The net mutual benefit of cooperation is what we call synergy. This concept is formally explored in the next section. The evaluation of synergy of a given FoS configuration constitutes the *static aspect* of the architecting framework.

5.2 The framework's static evaluation aspect: Synergy

We have seen the assessment of the value of a specific FoS configuration cannot stem from a global mission statement, overarching goals or requirements. Then, how do we evaluate FoS configurations? Is a given instance of a FoS desirable for participant systems? This is answered through the assessment of value against federation costs, from a local perspective.

5.2.1 Reasoning about synergy through utility and cost

To define and quantify synergy we will analyze the changes in the utility of federating for the participant systems. Assume a given system, has its own standalone utility and cost functions *f* and *g*, depending on its architectural variables x_1 as defined in Eq. 5-1 and Eq. 5-2.

$$U_1^{standalone} = f: x \rightarrow u \qquad \text{Eq. 5-1}$$

$$C_1^{standalone} = g: x \rightarrow C \qquad \text{Eq. 5-2}$$

The utility in Eq. 5-1 is defined in the same sense as in the literature section 4.2.3, as a measure of the value the system delivers to its stakeholders. All utilities in this chapter are defined in the range [0,1]. The costs of a specific system architecture, quantified in Eq. 5-2 are the full lifecycle costs including the development, manufacturing and operations of the system under examination. We assume those can be estimated from early stages using, for instance, parametric models (Wertz et al., 2011).

When system 1 cooperates with another system, called here *system 2*, with architectural variables x_2, changes in utility and cost functions of the system 1 appear, depending on the architectural variables of both systems involved $\boldsymbol{X}=(x_1 \cup x_2)$, and a set of parameters p (which, for instance, can include the extent and constraints of the cooperation). We assume we can model this changes via using functions f' and g' of $\boldsymbol{X}$. We will call the utility and cost experienced in the cooperation arrangement by system 1 $U_1^{cooperating}$ and $C_1^{cooperating}$. Eq. 5-3 and Eq. 5-4 describe this scenario.

$$\Delta U_1 = U_1^{cooperating} - U_1^{standalone} = f': \boldsymbol{X} \rightarrow \Delta U \qquad \text{Eq. 5-3}$$

$$\Delta C_1 = C_1^{cooperating} - C_1^{standalone} + Inc_{1\rightarrow 2} - Inc_{2\rightarrow 1} = g': \boldsymbol{X} \rightarrow \Delta C \qquad \text{Eq. 5-4}$$

Note in Eq. 5-4 the inclusion of incentives *Inc* for the cooperation. System 1 can receive or receive incentives from system 2. This terms enable the formulation here to capture economic transactions between systems to offset the costs for a peer and stimulate cooperation. To be able to reason about utility and costs of cooperation in a compact expression, we need to think about utility and cost in the same units. This is not unheard of in value-driven design techniques (Collopy, 2009). The best strategy to do this is to have a model of the function *f', call it h',* that assesses the economic utility of certain architecture, in the same units as costs. This could for instance represent a commercial revenue.

In these cases, the Net Benefit of Cooperating for system 1, NBC_1, can be obtained by simply subtracting Eq. 5-4 (the costs related to federation) from Eq. 5-3 (the benefits), as in Eq. 5-5.

$$NBC_1 = h'(\boldsymbol{X}) - \Delta C_1 \qquad \text{Eq. 5-5}$$

However this approach might not be possible for many systems, be ambiguous in case of scientific endeavours, or be subject to large uncertainty. Another way to consolidate the utility and cost measurements is to use the Pareto front as a bijective function to map utility to cost, in the native tradespace of the system, what we called SPF. Both approaches are very different, as the first compares utility and cost of architecture in the same units, while the other directly ***maps*** utility to cost.

The distinction between the two is somewhat analogue to the difference between *value-based* and *cost-based* pricing. On the following, we map utility to cost for a standalone system using Eq. 5-6 and Eq. 5-7:

$$f^{-1}: U_1^{standalone} \rightarrow x \qquad \text{Eq. 5-6}$$

$$C_1^{standalone} = g \circ f^{-1}\left(U_1^{standalone}\right) = F_1(U_1^{standalone}) \qquad \text{Eq. 5-7}$$

The function F_1 corresponds to the Pareto front in the system standalone utility-cost tradespace, the SPF. We assume the architect always chooses options from the Pareto front. In this manner, we can re-write Eq. 5-3 in terms of cost:

$$\Delta U_1{}' = F_1(U_1{}^{cooperating}) - F_1(U_1{}^{standalone}) \qquad \text{Eq. 5-8}$$

In Eq. 5-8 the term $F(U^{cooperating})$ maps the utility achieved by cooperating to cost by using the system standalone Pareto front. This means that it quantifies how much would the system have had to expend in order to achieve such utility by its own, standalone, means. Since now $\Delta U_1{}'$ is in cost units, we can subtract Eq. 5-4 and Eq. 5-8 for a single expression of benefits and costs of a specific cooperation between two systems. This is, as introduced above, the Net Benefit of Cooperation *NBC*, and we can write this for both system 1 and 2:

$$NBC_1 = \Delta U_1{}' - \Delta C_1 = F_1(U_1{}^{coop}) - F_1\left(U_1{}^{standalone}\right) - C_1{}^{coop} + C_1{}^{standalone} - Inc_{1\to 2} + Inc_{2\to 1} \qquad \text{Eq. 5-9}$$

$$NBC_2 = F_2(U_2{}^{coop}) - F_2\left(U_2{}^{standalone}\right) - C_2{}^{coop} + C_2{}^{standalone} - Inc_{2\to 1} + Inc_{1\to 2} \qquad \text{Eq. 5-10}$$

Note the incentives term difference in sign in Eq. 5-9 and Eq. 5-10 to denote the directionality of incentives. NBC is conceptually similar to Chattopadhyay's net perceived benefit explained in section 4.3.1, but, 1) the terms are rigorously consolidated in units through the use of the functions *F*, 2) incentives are exchanged between systems in the federation rather than used by global stakeholders. Now that we have the NBC of two systems delimited, we can write the generic expression for any system in the federation as in Eq. 5-11:

$$NBC_i = F_i(U_i{}^{coop}) - F_i\left(U_i{}^{standalone}\right) - C_i{}^{coop} + C_i{}^{stdalone} - \sum_j^N Inc_{i\to j} - \sum_i^N Inc_{j\to i} \qquad \text{Eq. 5-11}$$

Due to the mapping performed via the SPF, $F_i\left(U_i{}^{stdalone}\right)$ and $C_i{}^{standalone}$ have the same value and cancel out. Hence we can use:

$$NBC_i = F_i(U_i{}^{coop}) - C_i{}^{coop} + - \sum_j^N Inc_{i\to j} - \sum_i^N Inc_{j\to i} \qquad \text{Eq. 5-12}$$

The *NBC in* cooperation, depends, in the general case, on the architecture of all systems $\boldsymbol{X}$ and parameters *p*. According to our definition of FoS, the NBCs of all participating systems must be positive for them to join the federation. Moreover, due to their reciprocity, the conservation of incentives between systems holds in the overall FoS:

$$\sum_i^N \sum_{j,j\neq i}^N Inc_{i\to j} = 0 \qquad \text{Eq. 5-13}$$

Eq. 5-14 now defines synergy quantitatively. A synergy value is attached to a particular FoS configuration, and is the result of the NBC sum across participant systems:

$$Sy_{N\leftrightarrow N} = \sum_i^N NBC_i \qquad \text{Eq. 5-14}$$

Hence, synergy is the aggregation of the net benefit of cooperation through a FoS. Note that the sign of *Sy* does have any implication on the signs of each *NBC*. By the definition we proposed, only systems with positive NBC will be part of an FoS. However Eq. 5-14 remains valid to quantitatively evaluate speculative FoS configurations where some NBCs, or all, might be negative.

In the cases where no external incentives exist, computing synergy from Eq. 5-14 does not require us to establish an incentives scheme, due the conservation of incentives within FoS (Eq. 5-13). This corresponds to an absence of global stakeholders and external stimulation consequent with the principles of FoS and virtual SoS. Nonetheless, our formulation here allows us to identify cases of negative synergy, hence quantifying the amount of external stimulus that could bring the system to cooperate when internal compensations are not enough for FoS adoption. Furthermore, we can distinguish two cases depending on the need for incentives, which we name *strong synergy* and *weak synergy* conditions.

The formulation presented here makes no assumptions on the particular form of the functions that compute the NBC based upon a FoS architecture, hence general, non-linear synergies can be captured, and indeed, are expected. The limitations to this model of capturing synergy are directly related to the limitations of utility theory as means to capture all of stakeholder's needs and preferences, as discussed in section 5.6.

5.2.1.1 Strong synergy

We say a group of *N* systems experience *strong synergy* when the condition in Eq. 5-15 holds.

$$NBC_i > 0 \ \ \forall i \ when \ \ Inc_{i \to j} = 0 \ \ \ \forall i, j | (\ i \neq j) \qquad \text{Eq. 5-15}$$

The strict inequality denotes that we discard marginal scenarios for a federation to occur. In this case, $Sy_{N \leftrightarrow N}$ as per Eq. 5-14 will be semidefinite positive. In strong synergy conditions, incentives are possible, but even without them the NBCs are positive as Eq. 5-15 prescribes. This means that the utility and cost benefits of the systems participating are enough motivation on their own before any incentives.

By examining Eq. 5-12 we derive that for strong synergy to occur, all of systems have to experience any of these scenarios:

- An increase of utility at no additional costs;
- A reduction of costs with no decrease of utility;
- An increase of utility at a cost smaller than the corresponding cost for the same utility level achieved by their own means, that is, beating the standalone Pareto front;
- Both a reduction of both costs and increase in utility such that is beats the SPF.

In strong synergy all participant systems experience a performance that dominates their SPF. This can be compared to what economic literature calls *superadditive* synergy (section 4.1.4); that is, new value is created by cooperation that transcends the simple sharing of costs or amortization of investments.

We will graphically analyze this conditions in section 5.2.1 and see examples in the case studies. Strong synergy conditions are not always achievable in all FoS. However, that does not mean that we cannot find a rationale to federate through applying incentives between systems, as in the case of weak synergy.

5.2.1.2 Weak synergy

In the case of *weak* synergy, we recognize the cooperation might, per se, not be beneficial to some of the systems involved, which need be incentivized by others. The NBC of all participants is still kept above 0, by the exchange of incentives as Eq. 5-16 shows.

$$NBC_i > 0 \;\; \forall i \;\&\; Inc_{i \to j} \neq 0 \;\; \exists i,j |(i \neq j) \qquad \text{Eq. 5-16}$$

We can measure incentives in the same units as costs, as a monetary transaction. Incentivizing other systems represents a local cost for any given system, as expressed in Eq. 5-4. Hence systems will incentivize others only if the additional utility achieved beats such a cost. Conversely, systems will only accept to federate under incentives if those are larger than the costs incurred. This is represented by a positive NBC, after the application of incentives.

Let's clarify this formulation with an example, thinking only about the costs of federating, and leaving aside the additional system utility achieved; we will consider it constant in this illustration. Assume the existence of a system 1 and a system 2, such that system 1 supports system 2 and reduces its operation costs. Table 13 illustrates an arbitrary costs structure of cooperation and standalone operations for both systems.

Table 13. Example of synergy, NBC and arbitrary incentives for 2 systems.

System	$C^{standalone}$	$C^{cooperating}$	NBC before incentives	Slack for incentives	Incentive *Inc* exchanged	Final NBC
System 1	4 $	1 $	-1+4=3$	3$	-1.5 (1→2)	-1+4-1.5=1.5$
System 2	4 $	5 $	-5+4=-1$	-1$	+1.5 (1→2)	-5+4+1.5=0.5$
Synergy			***As slack***	**3-1=2$**	***As NBC sum***	**1.5+0.5=2$**

Table 13 describes a case where system 1 has a direct cost advantage of federating, of 3$. System 2, however, will have an increase of its costs of 1$. Note this implies a certain cooperating scheme between system 1 and system 2. This is easy to exemplify as some sort of operations that system 1 has allocated to 2 through an appropriate interface. Why should system 2 agree to such a cooperation scheme? Only if it receives incentives that overcome its cost increase. System 1 can allocate up to 3$ in incentives without losing its cost advantage completely. For illustrative purposes, assume system 1 allocates a 1.5$ incentive to system 2, that accepts to cooperate under this conditions. Therefore, the final NBC of federating for system 1 is 1.5$ and 0.5$ for system 2, making it desirable for both. Note we can quantify the scenario synergy as 2$ before having to propose any amounts for the incentives, as Table 13 hints.

5.2.1.3 Incentives and synergy

Note again, from the previous example, that we do not need to know the exact value of incentives before we can identify this case as favourable for federation, when synergy is positive.

The actual amount of incentives would follow from a negotiation between the systems, and is subject to dynamic changes through lifetime. Incentives will also be tied to specific timeframes or measurable resource trades involved. Fixing the actual value of incentives is out of scope of the current work, and is not needed to pursue the objectives here. Instead, quantifying the available slack for incentives, that is, synergy, effectively identifies the scenarios where federating is desirable, which in turn answers the research questions herewith. Moreover, in the cases of negative synergy, or no slack for incentives, the present approach allows to exactly quantify the external incentives required for early adoption of federations, when the systems do not exhibit positive synergy in early federation configurations.

5.2.1.4 Distributed and time-averaged Synergy

For our analysis it is useful to introduce here two additional quantities, the distributed, or system-averaged synergy, and lifetime integrated synergy. The former is simply defined as the division of the overall Synergy amongst the systems present, as in Eq. 5-17.

$$Sy_{distributed} = \frac{Sy_{N \leftrightarrow N}}{N} \qquad \text{Eq. 5-17}$$

$Sy_{distributed}$ gives a compact indication of the NBCs of the participant systems in the federation without needing to consider particulars. In a scenario where all systems experience a fair allocation of the overall synergy achieved in the federation, their NBCs would be equal to the distributed synergy. In addition to the systems averaging, we can also apply time averaging to the synergy value. Lifetime-integrated synergy is the time-average value of distributed synergy experienced by a system across its lifetime. This applies in cases when the system operates through multiple FoS configurations k, with duration D_k. In Eq. 5-18 we present the distributed and lifetime-integrated synergy that a system experiences through a FoS configuration path. This definition will be used for the dynamic aspects of the framework.

$$Sy_{distributed,integrated} = \sum_{k}^{M} \frac{Sy_k \cdot D_k \cdot \gamma(k)}{life \cdot N_k} \qquad \text{Eq. 5-18}$$

In Eq. 5-18, M is the amount of FoS configurations, and N_k is the amoung of system present in each configuration. The framework will also consider the usage discount factors γ, as we discuss in the case studies.

During the case studies we will point out when the results refer to overall synergy, or/and its distributed and lifetime integrated value. We will also present Sy as Return on Investment (RoI), dividing it by the cost of the interfaces needed to federate.

5.2.1.5 Extension to several utility dimensions

So far, for clarity, we have considered only a scalar utility to capture the local value of federating, without discussing the nature of the attributes composing it. In real-life, stakeholders might consider relevant several attributes and assign independent utility measures to them, dealing instead with an utility vector.

One option to deal with such scenarios is to use MAUT to consolidate the utilities for each attribute into a scalar, normalized value as discussed in the literature review. Another option is to generalize the treatment of this section to *N* utilities. The difficulties of such generalization condense to a single point: the need to map an arbitrary vector of utilities U to cost with a bijective function. That is, when using function *F*, we shall be able to derive a single cost coordinate from a vector of utilities.

Can we map the utilities U of cooperating to unique cost value regardless of the dimensions of U? This is indeed possible as F is a Pareto front, and this by definition establishes a bijective relation between its objective functions. Let's now briefly demonstrate that Pareto fronts are bijective. Take the front F from Eq. 5-7, and assume the tradespace of the system includes cost and this time two attributes, with independent utilities U_A and U_B.

$$C = F(U_A, U_B) \quad \text{Eq. 5-19}$$

Let us now sample the function F *twice* at the same arbitrary point (u_a, u_b), with the following results of the experiment:

$$C = F(u_a, u_b) \quad \text{Eq. 5-20}$$

$$C' = F(F(u_a, u_b))$$

If F is a functional description of a Pareto front, the vectors $\boldsymbol{J} = (C, u_a, u_b)$ and $\boldsymbol{J'} = (C', u_a, u_b)$ represent the merits of two different non-dominated architectures. Therefore, due the Pareto dominance principles explained in chapter 4 we have:

$$\boldsymbol{J} - \boldsymbol{J'} = C - C'$$

$$\boldsymbol{J} - \boldsymbol{J'} \textit{ non dominated} \rightarrow \exists i\, J_i' \leq J_i \quad \& \quad \exists i\, J_i' \geq J_i \quad \rightarrow C = C' \quad \text{Eq. 5-21}$$

In virtue of Eq. 5-21, any arbitrary set (u_a, u_b) can only map to a unique value of C and hence F is bijective. Additionally, this implies that F^{-1} exists and is also bijective.

5.2.2 Reasoning about synergy from a graphical perspective

Next, we attempt to describe and clarify the concept of synergy from a graphical perspective, via the tradespace plots of the utility and costs functions. As explained in section 4.2.1, conventional tradespace exploration represents a collection of alternative system architectures on a two (or more) dimensional space of metrics, providing an intuitive yet powerful tool for early stage decision-making. Figure 15 (left) shows a generic local system architecture tradespace of utility-cost, and the Pareto front.

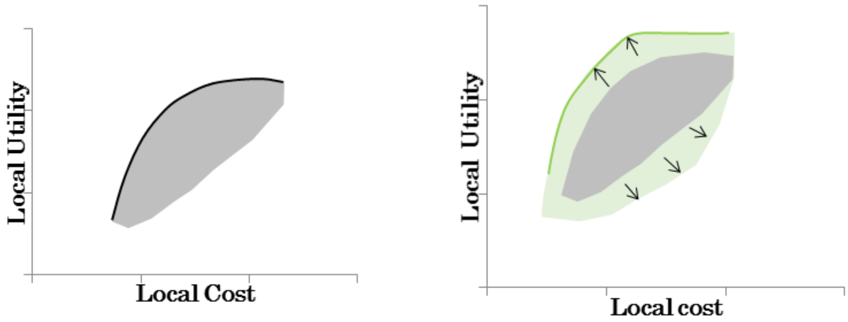

Figure 15. Left, generic utility-cost tradespace of a system which can be obtained with methods like MATE. All feasible architectures lie in the dark region, Pareto front in black solid line. Right, in light green, the general expansion of the feasibility region and new Pareto front due federation options.

As Figure 15 illustrates, adding federation options to the architectural trade-offs, will, in the general case, expand the feasibility region in all directions, potentially changing the position and shape of the Pareto front. Note that local utility perceptions of the federated options are based upon expectations of other systems' behavior.

Nevertheless, as we will see, if we expect other systems to have a rational behavior and also reason in utility-cost coordinates, the tradespace can be precisely laid out. We can now distinguish particular cases in the expansion of the feasible space and the displacement suffered by the Pareto front when federation options are added to a particular system. In a two-dimensional tradespace, all displacements can be codified by its magnitude and sign in two directions. It seems reasonable that federating has always an implementation cost associated –the cost of interoperability– but this does not necessarily lead to a net cost increase.

From a local perspective, displacements of an architecture to the left of the Pareto front, in the sense of cost reduction, are desirable. It is arguable if this may happen in an FoS, due the principle of operational independence. A cost reduction respect to the Pareto front can come only as a result of allocating local functionality and its related costs to another federate, in a way that is cheaper than the costs incurred by interoperability. A real life example would be choosing not to have any data storage capabilities in a computer and entirely relying in remote servers, when the latter is cheaper. Stored data is (still) critical to some of computer's functions; therefore this architectural decision would render it operationally dependent. However it may be possible to imagine scenarios where some capability has been installed in a system to maintain operational independence, yet the function it performs is allocated to a federate, and hence the operational costs associated are reduced. Figure 16 illustrates the discussion here.

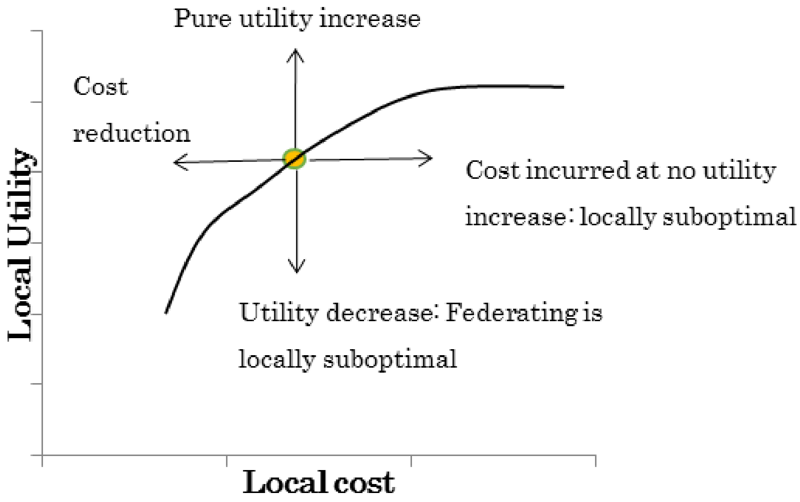

Figure 16. General directions for a Pareto front section displacement due federation options for a particular system.

The opposite case is incurring in costs, that is, being interoperable and cooperating with other federates at a local cost. This is a locally suboptimal decision if there are no utility increases associated. However, incentives can make it indifferent or even appealing to the system architect, as discussed in the previous section.

Shifting the Pareto front up means, in our depiction, an increase in local utility. The last option, displacements in the direction of local utility decrease, might occur if a system is supportive to another at the expense of underperforming its own mission. This is also a locally suboptimal decision which might enable weak synergy with other systems and could be compensated by incentives.

Note that all displacements in Figure 16 have an influence on other system's own local tradespaces, which now we introduce. Figure 17 shows a simplified view of the relations between two architectural tradespaces of systems which have an option to federate and the connection with the concept of synergy. When a federating option dominates the standalone Pareto front for both systems, this implies strong synergy. Weak synergy instead combines a suboptimal option for a system with a dominating option for the other. The former option can be made dominating also through incentives that can shift the cost coordinate of the option.

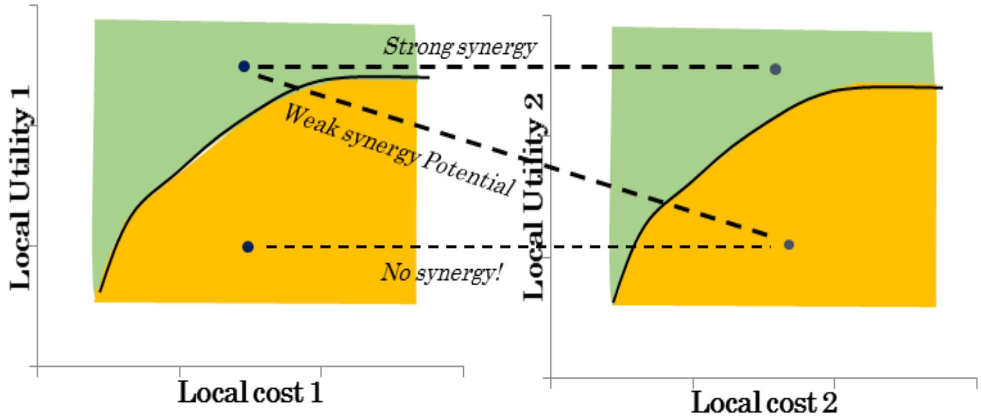

Figure 17. Left, the standalone Pareto front in the tradespace of a system (1), and outcomes of federating. Right, the corresponding points tradespace of another system (2) federated with the first.

Figure 17 shows a potential situation for weak synergy. It is only potential because it remains to be assessed if the conditions in terms of incentives for weak synergy apply. Would system 1 incentivize 2? How much incentive would 2 expect to adopt a suboptimal design point? Let's now examine in detail the conditions for synergy between two tradespaces.

Consider now both systems adopt respective architectures α and β such that they are federated. The reader can envision a practical case where system 2 is relaying data for system 1, enhancing the native bandwidth capabilities of the latter. Figure 18 depicts this situation.

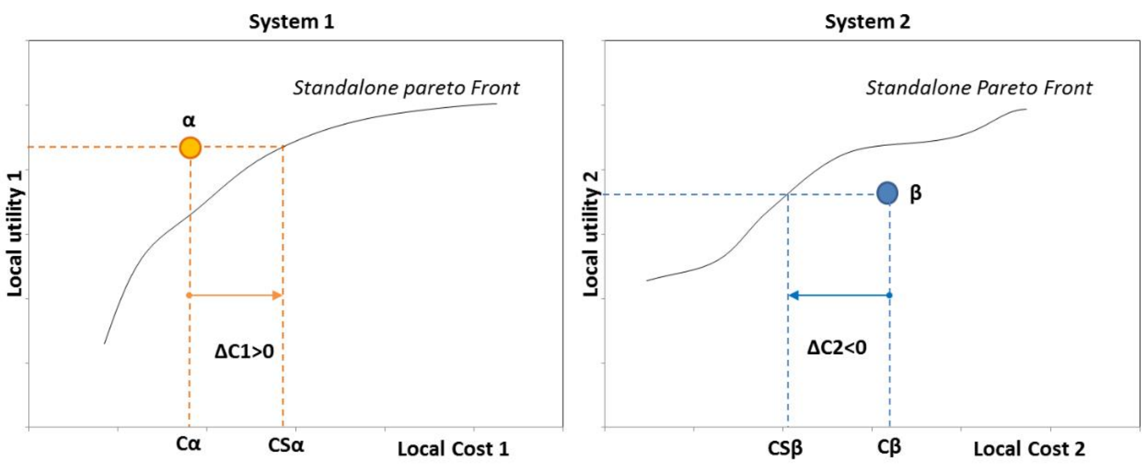

Figure 18. System 2 cooperating with system 1 under weak synergy conditions.

The result of the implementation of interfaces and the operational costs of cooperation C_β are such that the architecture β is locally sub-optimal. We can quantify the loss of optimality by comparing C_β to the cost in of achieving the same utility in the standalone Pareto front, CS_β. That reads as the costs incurred by system 2 to achieve the utility associated to option β by its own means, without cooperation. Conversely, system 1 achieves a better utility-cost coordinate by federating with system 1 than achievable as a standalone system, we can also quantify it as a cost variation ΔC_1 with respect to the cost of achieving such utility in a standalone fashion.

The next question that arises is if the systems would adopt architectures α and ß, and if that is beneficial for both. We argue that such is the case, marginally, when the cost advantage for system 1 is larger than the additional expenses incurred by system 2 (Eq. 5-22) and so the first can compensate the second for the incurred expense and still experience a cost advantage. This is equivalent to requiring both NBC's to be positive.

$$\Delta C1 > -\Delta C2 \rightarrow Sy_{[1,2]\leftrightarrow[1,2]} > 0 \qquad \text{Eq. 5-22}$$

Finally, consider the evaluation of the utilities and costs of local federated architecturesαand *ß*. This concerns the functions *f'* and *g'* defined in 5.2.1. Before we know the details of their cooperation and the extent of resources exchanged, and any other applicable parameters *p*, we cannot readily measure the architectural merits. While we might know the fixed capital cost associated with the adoption of the federated interfaces, we know little about the operational costs and the local benefit change; since those depend on the nature and extent of the cooperation with other systems, and vice-versa. Figure 19 illustrates this coupling problem.

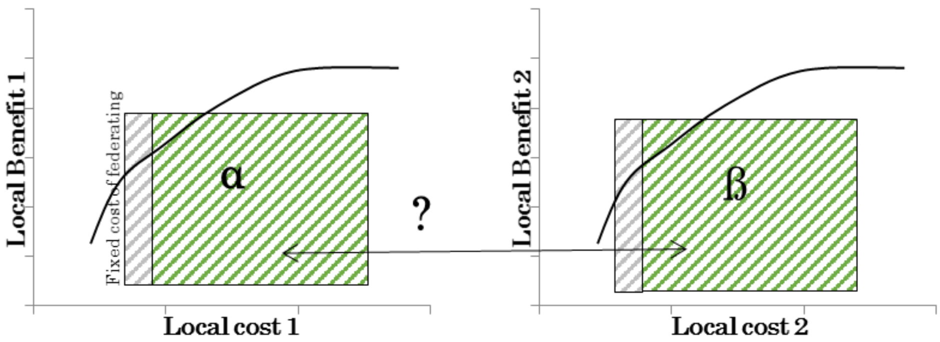

Figure 19. Federation option α in system 1, Federation option ß in system 2 and the copuling problem. The merits of the options are coupled and hence their positioning in the tradespace requires from a joint optimization.

In order then to solve this problem, we must optimize the architecture positioning in the tradespace depending on the cooperation details (amount of resources, time commitments, cooperation protocols) such that the NBCs of both systems are maximized. This optimization can be typically sometimes posed as a network flow problem, and/or heuristic search for the best resource allocation amongst systems.

The concept of synergy is powerful conceptual lever, as from the pool of potential FoS configurations, only the ones fulfilling weak or strong synergy conditions are interesting to the system architect(s). In any other case, one system will not decide to incentivize the other. Moreover, we can attempt to find the most synergetic combinations to architect the system of interest, which is the original motivation of the framework proposed here. After having defined the static aspect of the framework as the evaluation of synergy, we now need to examine the dynamic evaluation aspect.

5.3 The framework's dynamic evaluation aspect

Synergy evaluation has been introduced as the technique to assess a given FoS configuration. Nevertheless, FoS configuration shall necessarily vary and evolve as systems join and abandon the Federation. Moreover, we challenged ourselves to support studies of network externality effects and dynamics of the FoS.

During the history of a federation, the architect of every new system with federation potential ponders about the decision to join it based upon the effects of federating on their local tradespace, then adopts a certain design point and then system is comissioned. This creates a history of FoS configurations, or a path. Hence, the utility of the federated options is dynamic and depends on future, downstream FoS configurations. Figure 20 illustrates this process.

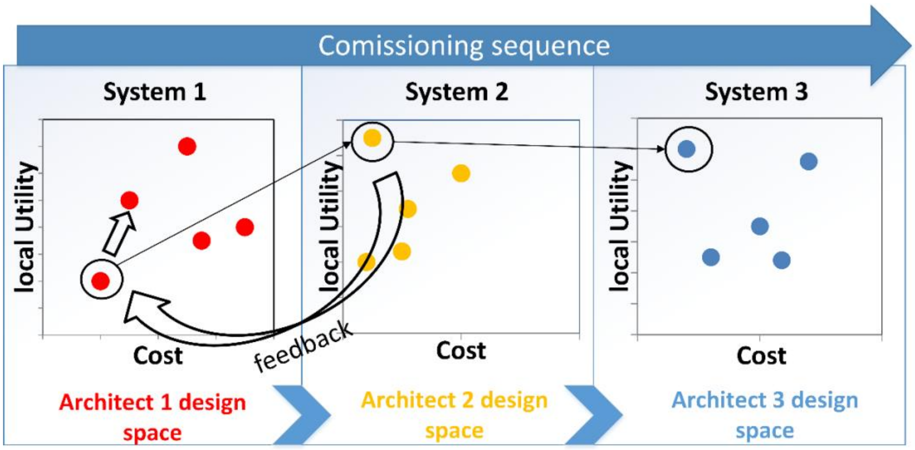

Figure 20. The dynamics of FoS: a succession of coupled architecting efforts.

Figure 20 also highlights notionally the changes in utility-cost coordinates of an architecture of an operational system, due to the commission of other systems during the former's lifetime. The general study of all the FoS architectures leads to a very large combinatorial space, of the order of A^N options, where N is the amount of systems present and A the architectural options for each system. In order to tackle this large combinatorial we will make use of the concepts of synergy and model each architect's decision in an MDP. This is detailed next.

5.4 The framework process

On the previous sections, we have defined the concept of synergy and proposed the means to measure it based upon to the utility and cost functions of the federated systems. These discussions partially cover the first research question on the nature of synergy in engineering systems (Q1). The dynamic aspect, brought up the question (Q2) about the appearance and evolution of a federation, is to be addressed by application of MDP principles. See chapter 4 for details.

The approach elements are consolidated here in a multi-step framework whose goals are 1) to characterize federation appearance and evolution as proposed, 2) to use this information to support the decision-making when architecting a system in a potential federated environment, which is a practical result of the research proposed.

The framework addresses the evolution of the federation through evaluating the alternative FoS configuration paths. At each time step, we consider a FoS architecture composed of existing systems and an additional new system *i*, which is the one being actively architected. The *A* architectural alternatives of system *i* generate *A* alternative FoS configurations. We then assess them all for synergy. Therefore, through the iteration, we analyze a particular system tradespace *i* including the synergetic effects with previously existing systems. The inclusion of this system to the FoS in one or several alternative architectures is considered. In the next step, we consider the alternatives of *i* effectively deployed and proceed to analyze system *i+1*. Step by step, a tree of evolving FoS architectures is built. Finally, we can feedback the information of synergy and NBCs upstream so that we find the chain of decisions (policies) that are optimal and include information of the future.

To address the effect of the uncertain future while maintaining the problem tractable, we use an MDP approach instead of transforming the architecting process into a stochastic optimization problem (Davendralingam and DeLaurentis, 2013). The general steps of the framework are depicted in Figure 21. On the following, we describe in more detail each of the proposed steps.

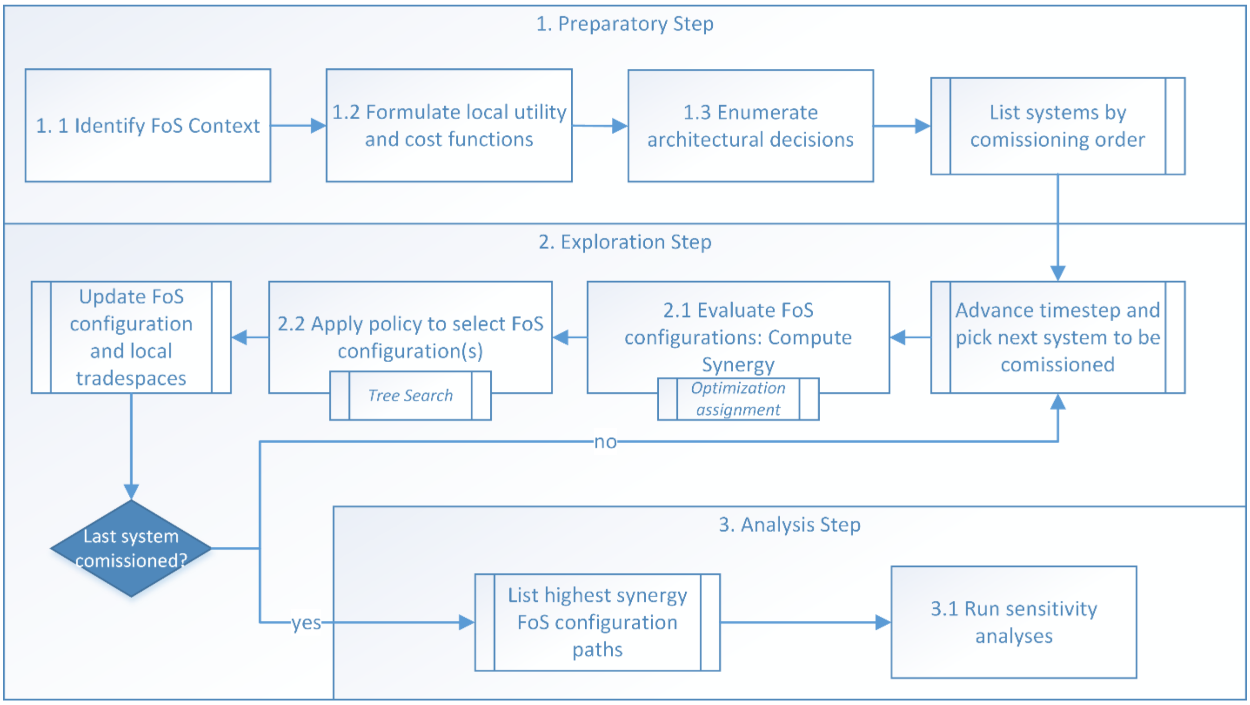

Figure 21. Proposed framework.

5.4.1 Preparatory step

The preparatory step defines the framework application domain and system properties before delving into the dynamic exploration of the federation evolution.

Step 1.1: Identify FoS context

At the start of the analysis, the FoS does not effectively exist. The FoS context has to capture then the technologies, potential future systems and federation boundaries, if any, of the domain under consideration, so that all the relevant information to assess the emergence and evolution of the federation is taken into consideration.

Formally, we define a FoS context as a set of exogenous and endogenous factors. The endogenous factors include the definition of the class of systems to consider for federation, the specific technological forms, the capabilities or resources that could be exchanged, the system functions, the systems individual goals, their conventional concept of operations, the potential technical processes supporting cooperation, the system's commissioning cadence, and the analysis time horizon. The exogenous factors include any bounds or external effects on federation growth, of topological and regulatory nature.

An example of FoS context would be the Low Earth Orbit (LEO) based EO, including as potential federates future satellites that 1) could embark FoS interfaces, 2) have no managerial or operational barriers to join FoS. A detailed discussion of this case is found in (Lluch and Golkar, 2015). Another FoS context could be the automotive park of vehicles in a city with mobile connectivity.

We must also consider the difference between a system architecture and the actual amount of instances of the system deployed. In the case of FSS, each satellite tends to be a custom development and only one or a few of identical satellites are deployed. However, in case of V2V or other FoS, tens of thousands of instances of particular system architecture might be deployed. This has to be accounted in the specifics of the framework's application.

Finally, a scope in time horizon for the analysis is also needed. The analysis is conceptually infinite. The opportunity for an FoS to occur or collapse is never-ending provided the technological forms and the engineering field we are considering are still valid. In practice, events in remote futures further than the lifecycle of the systems being engineered in the present are of limited interest for the system architect, and this is duly captured by the discount factors of the MDP process (see section 4.4.2). The time horizon for the analysis needs to span at least a lifecycle of the first systems adopting the federation. Typically this will represent scenarios of 10 years.

Step 1.2: Formulate local utility and cost functions

In this step, we characterize systems in their local tradespaces, identifying the key metrics of interest for systems stakeholders and modelling their preferences as utility-performance functions. We also gather the necessary elements to later estimate the cost of the systems as functions of architectural decisions.

Step 1.3: Enumerate architectural decisions

In this step, we enumerate the different architectural options available to system architects from a set of decision variables. We both list non-federated and federated alternatives, hence characterizing the cooperation interfaces and their potential capabilities and costs.

From the non-federated architectural options set we can derive, with an adequate system model, the standalone tradespace of the system and obtain the SPF. In this manner, we make available the F function. As introduced in section 5.2.1, in the cases where an economic return is directly available from each architecture via a function $h'(x)$, we do not require to generate the SPF. The federated options are analyzed in the exploratory step for each system in the context of synergy computation, and requires as inputs the architecture of other systems too.

The models used to evaluate the utility and cost of each architecture, both at the preparatory and exploration steps, do not require to include all aspects of the system, but only those related to the considered system performances, architectural decisions and resources to be traded. That is, if FSS cooperation is based upon a data operand, the model includes the aspects of data exchange (communications equipment models, ground stations, ISL subsystems) but does not consider for instance detailed modelling of other spacecraft subsystems, like thermal or propulsion. However they might be affected by the federated interfaces as a secondary effect. The specifics of this are discussed in the case study and in previous work (Ignasi Lluch i Cruz and Alessandro Golkar, 2014).

5.4.2 Exploratory step

This is the key step of the approach, which iteratively builds an MDP graph as the future systems listed in step 1 are sequentially architected. Each system alternative leads to a different FoS configuration with a given synergy state and local NBCs. Figure 22 reflects this idea.

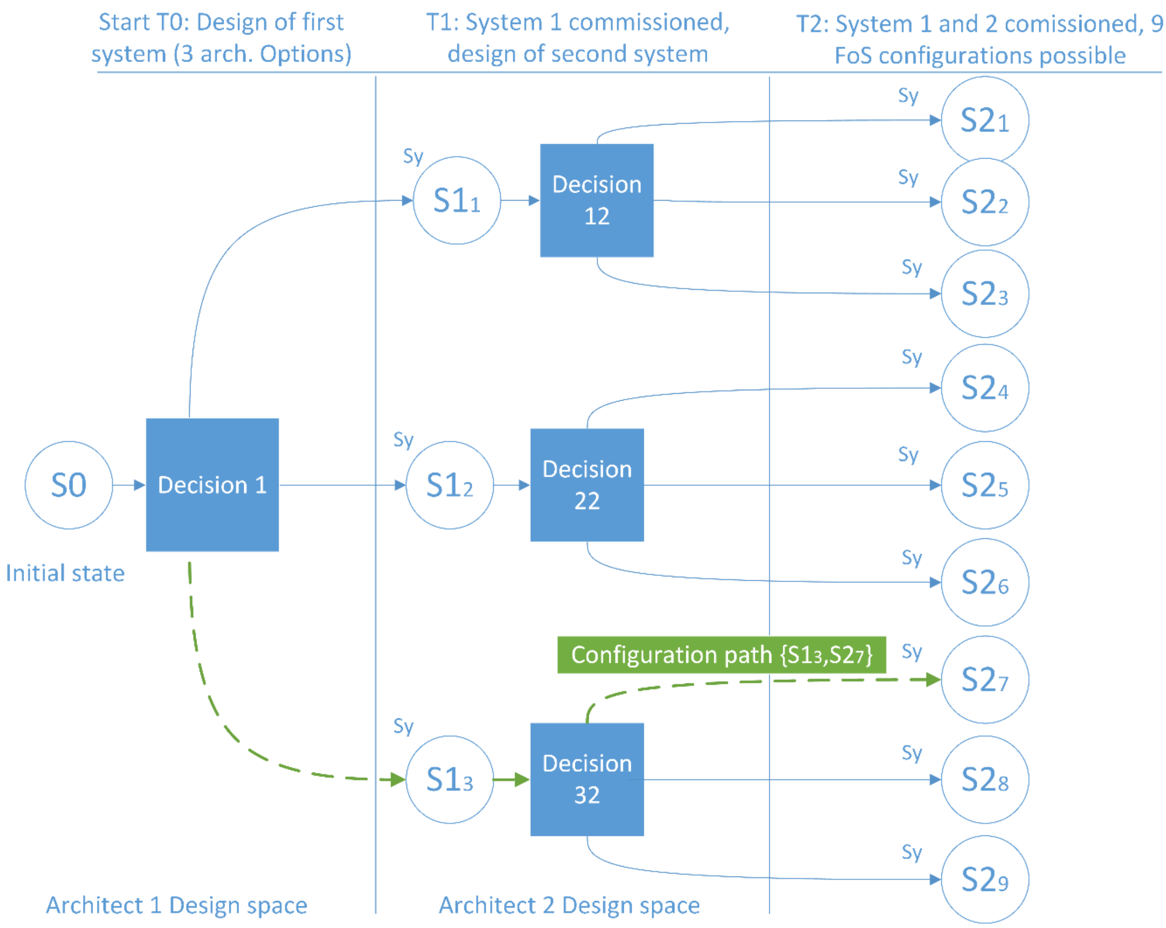

Figure 22. The FoS architecting problem is a Tree with FoS Configurations as nodes, and decisions that stem from an expectation of future rewards. This constitutes a Markov Decision Process (MDP). In this example, 2 systems with 3 architectural options each are commissioned. Find highlighted the definition of a specific configuration path.

The graph in Figure 22 considers the addition of two systems with 3 architectural alternatives each. When the first system is added in T1, the FoS architecture, represented by Markov states, can transition to $S1_1$, $S1_2$ and $S1_3$, which are composed of the previous systems plus one of the specific architectural alternatives of system 1, respecting the Markov property. Due to the particulars of our architecting problem, the MDP graph has a tree structure without loops (i.e, it is not possible to go back in time). Each state is assessed in terms of synergy. While only positive synergy states are self-sustainable for a federation, negative states might be acceptable if they grant FoS transitions to higher synergy. The MDP action node represents the architecting decision for each system.

The decision 1 in the blue box, Figure 22, is made by the architect of system 1, selecting which of system 1 alternatives is more convenient. We used above the q^* notation of MDP's to denote the expectation on rewards of each decision. We here assimilate the MDP reward to the discounted and distributed synergy of a state. The system architect then would choose an option that picks the maximum rewards downstream, *max(q*).* An interesting discussion is that this is similar to an adversarial search problem (Russell and Norvig, 2002), since the next decision in the graph is to be taken by architect 2 , who does only capture the rewards from T2 onwards.

Since this is not an adversarial but a cooperative situation, and both systems pursue the highest synergy as this is mutually beneficial, arguably they would take the same decision. However, the subtle change on perspectives enables the existence of configuration paths with decreasing positive synergy. This is the case in some of the case studies, and illustrates the potential mechanism for collapse of the federations. We study this in more detail with the applied examples of the cases studies.

Finally, note that while general MDPs have transition probabilities, the probabilities of all transitions in this MDP are assumed $P=1$ as the outcome of selecting an architectural alternative is not stochastic.

Step 2.1: Evaluate FoS configurations: compute synergy

Each state, or FoS configuration, is evaluated for synergy using a federation model, which outputs costs and utilities for all federates depending on their architecture. Since the cooperation scheme (allocation of resources) that captures maximum synergy might not be trivial, we may need to run suitable optimization methods to allocate the federation resources, using the toolset described in chapter 4. Then we are in condition to compare optimal allocations to optimal allocations.

In problems where the time of commissioning between systems is comparable to the lifecycles involved –such as satellites– we need to introduce a discount factor, a common feature in MDP literature, to account for the relative preference of the architects for rewards closer in time. Hence each system sees the synergy rewards of future FoS configurations corrected with the closeness to them.

In most cases, the potential FoS configurations and hence MDP states is very large. For instance, for a case with $N=10$ systems and $A=12$ architectural options per system we have more than $6.7 \cdot 10^{10}$ states, estimated as shown in Eq. 5-23. For reference, if each state could be computed in CPU time of 0.01 s, 12 CPUs would need 651 days to compute the full MDP state tree. Note state synergy computation might require an optimization step. Therefore, we only analyze states in promising paths, using search methods or fixed policies in step 2.2.

$$M = \sum_i^N A^i \ (8) \qquad \text{Eq. 5-23}$$

Step 2.2 Apply policy to select FoS configurations

In principle, once the full MDP tree is built, we can work backwards to compute the q^* values of each decision, and find the optimal policies at each step, using the classic recursions available in the literature to solve MDPs. Hence, we can determine the best architectural decision for each system, and the outcomes of these decisions. This supports the local decision-maker or the architect to understand if a particular system should be architected for a federation, and how can a particular system influence the evolution of the whole FoS.

Nonetheless, *a priori*, it is computationally unfeasible to compute the states for the full tree in realistic cases due to its dimensions. This framework proposes two options to face this problem, which is a classic combinatorial explosion. The first option is to use tree search mechanisms, including alpha-beta pruning (Russell and Norvig, 2002) and greedy algorithms, so that we exploit the time structure of our problem, using the synergy of the next step as a *greedy* estimator of future attainable synergy. This works well in MDPs including discounts. In some situations, if the pruning mechanisms are effective enough, it is possible to explore the full tree remainder and obtain the global optimal set of systems' decisions.

The second option is to assess a limited set of MDP policies, based on heuristics, to generate a limited set of FoS configuration paths. If the system's capabilities are comparable, adding new systems, especially at later stages of the federation, cannot create drastic changes in the synergy trends observed. This well-behaved synergy evolution allows for the generation of good heuristics. We illustrate the applications of both these approaches in the case studies.

5.4.3 Analysis step

This step examines the results of the best FoS configuration path(s) obtained, via policy or tree-search approaches. With this information, we attempt to identify emergence, trends, plateaus, and dead-ends for the FoS evolution, what is the return for participant systems of federation, and what technical parameters drive it. We use the metrics of overall synergy, distributed and lifetime-integrated synergy, and also the federating RoI measure in the cases interface adoption costs are comparable to the returns.

5.4.4 Run sensitivity analyses

This step includes obtaining new FoS configuration path(s) with different assumptions and policies. We will compare the synergy results typically for different FoS topologies, local stakeholder utility functions, interface cost and technical capabilities, and any applicable case specifics.

5.5 On coordination mechanisms, boundaries and authority in federations

As captured in the working definition, we recognize in FoS a voluntary facet. Systems can join or abandon the federation on their own initiative, that is, no system can be coerced to remain or cooperate in detriment of its own benefit. As announced by the federalism principle of subsidiarity, the authority remains on the particular components instead of an over-arching authority.

However, the systems might need to establish amongst themselves some FoS governance or coordination mechanism. On one side of the spectrum we can have an FoS based on a collection of ad-hoc, peer-to-peer arrangements with little coordination or oversight, and on the other, closed communities with managed cooperation. Figure 23 captures this range.

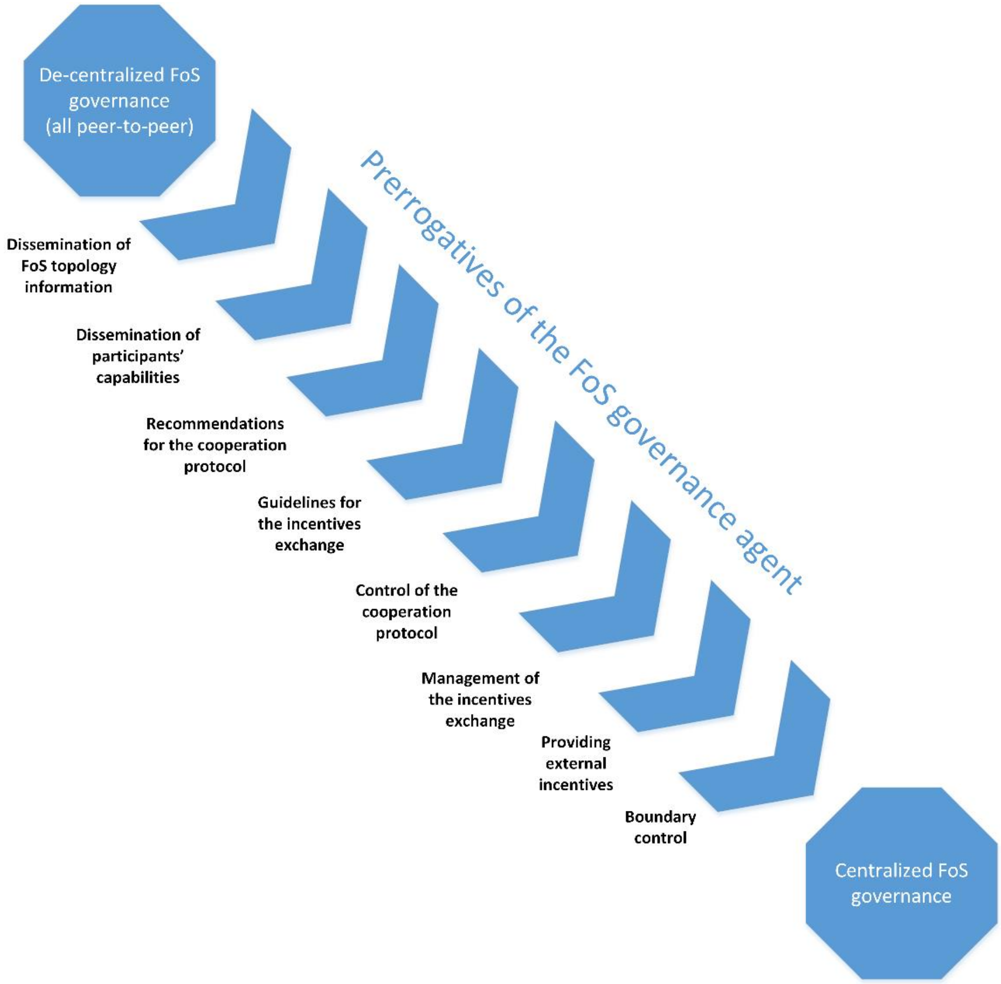

Figure 23. Transition from peer-to-peer regulated FoS to centralized governance, by progressive allocation of roles to an FoS governance agent.

The governance duties can be performed by a panel of representatives of the participant systems' stakeholders, or by a third-party that fostered the federation and has a key leverage on the cooperation process (for instance by controlling the technical standard). The coordination mechanisms in FoS range from purely informative processes (dissemination of topologies, listing of participant systems capabilities) to decision making on cooperation protocols, incentives management, and acceptance of new systems in the federation, that is, managing FoS boundaries.

Nonetheless, if the FoS experiences negative synergy per se but operates in virtue of an external incentives scheme, the boundaries with directed SoS become more blurred, as an external incentivizing agent might be able to impose conditions on the cooperation. The same type of leverage exists when the exchange of incentives or cooperation protocol is controlled by a third-party, as is the case in the sharing economy. We attempt here to list all the attributions that can be allocated to the FoS governance agent:

1. Dissemination of FoS topology information (when applicable to networks), and existing participants.
2. Dissemination of participant's capabilities and needs.
3. Elaboration of recommendations for the cooperation protocol (interface standards).
4. Elaboration of guidelines for the incentives exchange.
5. Control of the cooperation protocol.
6. Management of the incentives exchange.
7. Providing external incentives to systems.
8. Boundary control –Deciding on acceptance and expelling of systems from the FoS–.

Figure 23 introduced this elements as steps in a progressive shift from self-organizing FoS to closely managed FoS, which are in practice very close to directed SoS. The different case studies in this thesis describe FoS at different steps.

5.6 Assumptions and limitations

The main limitations and assumptions of the approach introduced here are the *non-simultaneity of architecting, the usage of the Pareto front as bijective utility-cost function, the indifference to architectural alternatives, and the rationality of system architects.*

The non-simultaneity of the architecting of different systems means that, from the perspective of the Nth system being architected in the FoS, we assume the architecture of previous systems and future ones is not being devised at the same moment. While we accept that future system architectures are uncertain, and might be reactive to present architectural efforts, we also assume that the systems are deployed one at a time. This appears like a reasonable assumption for most FoS, and makes the decisions taken by different architect's non-concurrent. If this would not be the case, game theory approaches instead of MDP would be more adequate (Von Neumann and Morgenstern, 2007).

The usage of the Pareto front as bijective utility-cost function was discussed at the beginning of this Chapter. This assumption enables to compare local utility changes across tradespaces, especially when the systems performances are not directly connected to economic revenue. However, when available, a model of the economic returns of each system, against costs, would be preferred.

The indifference to architectural alternatives means that system architects have no a-priori preference between non-dominated Pareto solutions. For instance, they will adopt high cost & high utility federated solutions if those dominate the standalone ones. This assumes a generic case without specific design constraints: shall those exist, they can be added to the tradespace analysis.

The rationality of systems architects means that if there is a net cost advantage of federating, the corresponding federated architecture will be adopted. However, federating may include some risks that would require the payoff to be larger than a certain threshold (Chattopadhyay, 2009). Such threshold can be included in the cooperation costs term.

5.7 Application to case studies

The proposed methodology will be applied to a case study in Satellite Federations for EO, a retrospective study on ridesharing and a case study on peer-to-peer community wireless networks. These cases differ in technologies, typology of the systems and stakeholders involved, degree of deployment in the real world, and FoS governance mechanisms.

The first case study fulfils the major applied motivation for this thesis work. Besides the particular technical insights it derives, its role in the wider academic context is to consolidate, anchor and foster the FSS research efforts started in 2013. Such research efforts span today a growing academic community involving more than 4 universities, and spur industrial interest. This case study is forward-looking and of predictive nature as FSS has not been widely adopted by satellite manufacturers and owners.

The second case study is focused on one of the flagships of the sharing economy. We will use real data on a concrete, existing case of sharing economy/FoS, and hence perform a retrospective analysis of this case. Applying the framework outside the engineering scope when stress-test its processes and theoretical fundaments, and examine in closer detail the similarities and differences between engineering FoS and sharing economy markets. This case is intended to yield methodological validation for the proposed framework. Finally, the third case study will analyze the FoS emergence and dynamics in rural WCN. This is a case study technologically similar to FSS but exhibiting different network structure and systems' relations.

Chapter 6 Satellite federations case study

This chapter introduces a case study in federations of satellite systems, the main background motivating this thesis. For details on FSS, the reader is referred to Chapter 4, which includes a taxonomy on distributed space systems, the specifics of FSS and a review on relevant networking technologies.

This chapter demonstrates the framework point by point, in a realistic case composed of five satellite systems. This nominal case is then followed then by parameter sweeps on discount factors, number of systems, stakeholder utility and cooperation constraints. With this body of results, we derive insights for the realization of federation.

6.1 Framework implementation on FSS: nominal case study

This section details the implementation of the approach with an illustrative satellite federation case, following the steps illustrated previously in chapter 3. The case introduced here acts as a nominal case from which several alternative cases will be derived.

6.1.1 Identify FoS context

In this case study we analyze the deployment of five Earth Observation (EO) satellite missions in Low Earth Orbit (LEO). This context has been proposed before (Lluch and Golkar, 2015) as a relevant case for FSS. Due the short duration of ground station contacts in LEO, and the large amounts of data generated by the new generation remote sensing instruments, downlink datarates are being pushed to their theoretical limits (Rosello et al., 2012). A federation among LEO missions could alleviate this problem, creating a pool of bandwidth and contact opportunities for federates in need to downlink larger data volumes. In this specific context, the research question is formulated as follows: does it add value, for the first mission and subsequent, to add the necessary interfaces for cooperation, embodied by ISLs? Under what cost assumptions will the federation thrive? How does FSS compare with alternative space-to-ground communication architectures?

The five missions will be generic EO spacecraft, with parameters modelled after typical EO missions and commissioned at yearly cadence.

Table 14 includes the orbital parameters and lifecycle expectation of said missions, all corresponding to Sun-Synchronous orbits (SSO), taken from a randomized distribution as in (Lluch et al., 2015). We do not consider any previous mission with interfacing capabilities to have been deployed previous to the first system commissioning.

Table 14. Orbital parameters of potential federate set. The parameters are Semi-major axis (SMA), eccentricity (Ecc), inclination (Inc), Argument of Perigee (AP), Right Ascension of Ascending node (RAAN), initial mean anomaly (MA_0), Beginning Of Life (BOL) and End of Life (EOL).

ID	**SMA (km)**	**Ecc**	**Inc (deg)**	**AP (deg)**	**RAAN (deg)**	**MA_0 (deg)**	**BOL (yr)**	**EOL (yr)**
Sat1	7049583.76	0.000089	98.073	110.5837	278.5846	163.8229	1	11
Sat2	7076612.5	0.001508	98.183	102.3389	212.9442	158.5433	2	12
Sat3	6834274.84	0.000913	97.238	114.1947	2.761352	172.4856	3	13
Sat4	6995764.74	0.000378	97.858	111.5622	231.151	260.1792	4	14
Sat5	7146445.76	0.000846	98.471	110.5294	299.232	275.5554	5	15

6.1.2 Formulate local utility and cost functions

The two key performance metrics that will be considered here are downlink bandwidth and communications latency of each federated mission. The downlink bandwidth is measured in Gbit/Orbit of mission data downloaded per orbit. This is taken as an averaged value amongst several orbital revolutions since ground and inter-satellite link contact periods change at each orbit. The communications latency is defined as the time between data generation on board the spacecraft and its reception on the ground. All through this chapter, latency is measured in minutes. Typical LEO missions with one or two ground stations experience a mission data latency between 90 and 30 minutes (Lluch et al., 2015).

Performance-utility measures

The performance metrics that will be used to evaluate the architectures are bandwidth and communications latency. The link between architectural decisions and performance metrics is established by the model detailed in forthcoming section. The bandwidth downlink volume in Gbit/orbit and the average latency in minutes have been mapped to utility through the piece-wise linear functions plotted in Figure 24 and Figure 25. This mapping captures the fact that the stakeholder's utility function exhibits a varying response across the performance range.

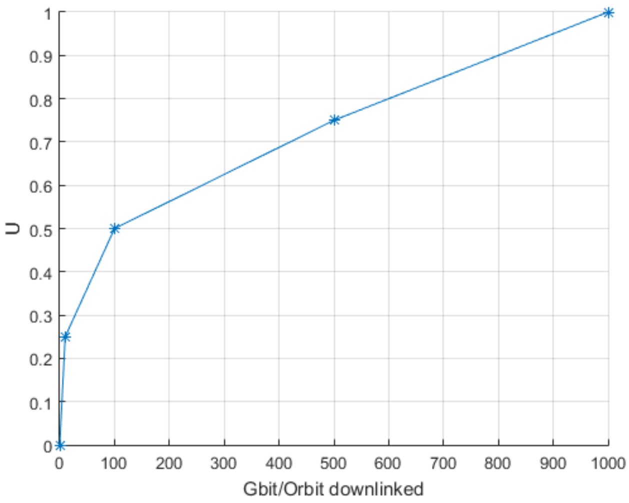

Figure 24. Bandwidth downlink volume mapped to utility, by using a 4 pieces linear function to simulate non-linear, concave behavior. Downlink volume performances above range are assigned utility 1. Utility 0.5 has been assigned to a 100 Gbit/Orbit volume. For reference, this can be achieved with a 160 Mbit/s link in a 10 minutes ground station pass, a routine LEO performance.

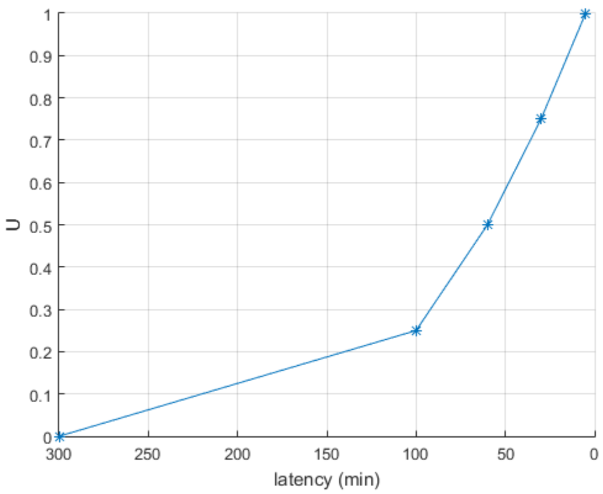

Figure 25. Latency performance mapped to utility, by using a 4 pieces linear function to simulate non-linear, convex behavior. Latencies above 300 minutes are assigned utility 0, and latencies below 5 minutes are assigned utility 1. Typical LEO latencies are between 30 and 90 minutes.

The nominal case uses this arbitrary albeit realistic utility mappings. For bandwidth downlink, the mapping starts from 0 Gbit/orbit and assigns utility 1 to 1000 Gbit/orbit. This is equivalent to the state of the art in data volumes downlink per orbit. 1000 Gbit/orbit corresponds to 20 min ground station access per orbit with a rate of 800 Mbit/s. Those are roughly the performances of the WorldView-4 imaging satellite ("Eoportal missions directory," 2016). Similarly, for latency, 5 minutes or less, which is fundamentally real-time for a space segment, we have assigned utility 1. Latencies of more than an orbit pass –above 100 minutes– have been assigned 0.25 utility. Latencies of 300 minutes and more have been assigned utility 0, as the less capable of ground segments envisioned in Table 15 delivers better performances (Lluch et al., 2015).

The fact that not all missions may have same utility measures, and that the shapes for their utility functions might not be as depicted above, is explored in Section 6.4.1.

6.1.3 Enumerate architectural decisions

These two performance metrics depend on the communications architecture of the mission, including the amount and location of ground stations used, the capabilities of the communications subsystem onboard the spacecraft, the usage of additional geostationary relays (Frank J. Stocklin et al., 2012), and of course, the adoption or not of the required equipment for federation. We then formulate a design space of four decision variables: Ground Segment Architecture options (GSA), Geostationary Relay (GR) options, Space to Ground Rate (SGR) options, and Federation Options (FO). All associated costs herewith are in USD units, referenced to fiscal year 2016. All costs are intended to represent lifecycle costs. Due the nature of each of the decisions and their cost structure, for some decisions the cost we use is a fixed expense (like for SGR), the recurrent cost integrated through lifetime (like to GR), or a combination of fixed and recurrent (for GSA).

Ground Segment architecture options

In order to represent the diversity of ground segments possible, up to six options are included under this decision, including building dedicated stations, commercial lease, and governmental lease. Table 15 summarizes these options. The costs include recurrent costs (maintenance or lease) and fixed costs when applicable.

Table 15. Options for GSA and its inputs for the cost model. Options 1 and 2 use models, while 3,4 and 5 are based upon real data.

Option	Descriptor	Ground stations	Costs associated	Observations
0	No GS	-	-	Requires use of either FO or GR
1	Dedicated, 1 station	Located as Svalvard	15 MUSD for commissioning plus maintenance (see annex II)	
2	Dedicated, 2 stations	Located as Svalbard and Troll	25 MUSD for commissioning plus maintenance (see annex II)	
3	Commercial lease, per pass	Alaska NP, Hawai and Dongara locations modelled after SSC as in (Harris et al., 2016)	Charges 250 USD per satellite pass (Harris et al., 2016)	150 Mbit/s rate limit
4	Commercial lease, fixed rate	Seattle, Alaska and new Zealand locations modelled after SpaceFlight Networks	Charges 50kUSD monthly (Spaceflight Networks, 2017)	300 Mbit/s limit
5	Extensive lease, governmental	6 stations, Modelled after NASA NEN (Harris et al., 2016)	Charges 450 USD per satellite pass (NASA, 2015a)	

Geostationary relay options

The geostationary relay (GR) options are modeled after NASA's TDRSS (Frank J. Stocklin et al., 2012) as shown in Table 16. For bandwidth and latency estimation purposes, the missions making use of GR are assumed to make use of it five minutes out of every 20 minutes. This is a very frequent, operational usage of GR, that very few missions engage in routinely due its high costs per minute (NASA, 2015b). Such usage profile amounts to 25% of orbit time in contact, which is comparable to ground station contact for the GSA architectures with ID 2,3, and 5 shown in Table 5. The high costs of GR services stem from the amortization costs of the expensive space and ground segments required, and the limited amount of users. Furthermore, GR systems are constrained in the amount of users they can serve, hindering positive economies of scale. Therefore, routing most mission data through GR has been deemed in previous studies as an ineffective practice (Eilertsen, 2012). Most missions route only telemetry or command data via GR, or only use it as backup or contingency mode. Notwithstanding this common wisdom, comparing quantitatively the operational use of GR against other ground segment options and FoS is part of the insights that can be obtained in this demonstrative case.

Table 16. Geostationary relay options, real costs taken from (NASA, 2015b).

OptionID	Descriptor	Costs associated	Observations
1	Not to use GR	-	Requires use of either FO or ground stations
2	Multiple Access mode	12 USD per minute	Up to 3 Mbit/s
3	Single access mode	120 USD per minute	Up to 300 Mbit/s

Two modes are available for the GR: Multiple access (MA) in the media is shared amongst several spacecraft and lower data rates are supported, and Single Access (SA) which allocates a dedicated high-speed link to the customer mission.

Space to ground rate options

The costs of the architectures are dominated by the GSA strategy, however characteristic costs for the SGR have been included and are part of the model. The SGR option performance and cost is a product of the spacecraft communications subsystem and on-board data handling responsible for downlinking spacecraft mission data. For reference, note that a COTS communications subsystem in X-band for small satellites capable of 150 Mbit/s is about 1 MUSD (Surrey Satellite Technology US, 2017). The USCM8 model (Wertz et al., 2011) proposes figures between 2 and 37 MUSD when communications subsystem mass is between 10 and 200 kg.

Rates of 500 Mbit/s and above are less frequent and have been priced at 30 MUSD. These costs are fixed, and the corresponding recurrent costs of operating these transmitter are allocated to the GSA recurrent cost component. The Space to Ground (SGR) options are listed in Table 17.

Table 17. Space to Ground Rate Options. For details on estimations see Annex II.

OptionID	Descriptor	Costs associated	Observations
1	No S-G communications equipment	0	Requires use of either FO or GR
2	50 Mbit/s transmitter	0.5 MUSD	
3	150 Mbit/s transmitter	2 MUSD	
4	300 Mbit/s transmitter	10 MUSD	
5	500 Mbit/s (Kankaku et al., 2013)	30 MUSD	Possible using several channels

Federation options

For the nominal case we consider only two options, either not to include the interfaces for federation, or to include them in the design of the participating spacecraft. The design impacts of including such equipment on the rest of the spacecraft subsystems, specifically on power and attitude control, have been explored before (Lluch and Golkar, 2015) and are accounted for in the cost budget. The cost here is a lifecycle integrated cost (including fixed and recurring federation costs). This figure is swept from 1 to 24 MUSD in the nominal case.

Table 18. Federation Options.

OptionID	Descriptor	Costs associated	Observations
1	No FSS equipment	0	Requires use of either Ground stations or GR
2	FSS equipment	Variable	

Architectural enumeration

The exhaustive combinatorial enumeration of all options introduced above yields a tradespace of 180 options, of which only 75 are feasible, give the need of have some form of space-to-ground communication. Simultaneous use of GR and FSS has been also ruled out as these options are a mutual substitute. Table 19 summarizes the options for the four decision variables chosen.

Table 19. Summary Architectural decisions and options for the potential federates.

GS arch options (6)	S-G rate options (5)	GR options (3)	FSS options (2)
{0,1,2,3,4,5}	{0,50,150,300,500} Mbit/s	{no use, MA, SA}	{no use, use}

6.1.3.1 Performances modelling

A model for latency, a model for bandwidth, a model for cost and a satellite orbit propagation backbone enable the architectural evaluation. The latter is based upon Paul Grogan's FSS toolkit (Grogan et al., 2014). The tool computes the geometry and visibility amongst spacecraft and between spacecraft and the ground. Taking into account the amount of necessary evaluations of the architecting process, it is of utmost importance to use lightweight models with adequate balance between execution speed and fidelity.

Based upon on the orbital propagation, all passes and contact times between assets can be derived. This supports the cost estimation, performed with a bottom-up model based upon the data in Table 15 to Table 18. The maintenance cost model for the ground stations follows the recommendations in (Wertz et al., 2011).

Latency for each mission is sampled at discrete steps through an orbital simulation that includes all federates and their assigned ground stations as per architectural definition. At each sampling time step, latency is computed as time interval to the next *'connected'* time-step in the simulation. A spacecraft is in a 'connected' state if there is visibility to a ground station, either direct or through a one-hop relay via a federated spacecraft. Both the source and the relay spacecraft need to have FSS interfaces in the second case. The one-hop limitation is enforced to emulate network congestion limits. Moreover it has been shown that most of FSS relays in LEO can be accomplished with just one hop (Lluch et al., 2015).

Notably, each spacecraft has its own ground station(s) assignment depending on the architectural definition, and they cannot directly downlink data to other ground stations. They can do so indirectly, by using another federate as a relay. This is intuitively one of the key advantages of federating: the possibility of accessing to an extended pool of ground stations with a single interface. Shortly we will quantify such advantages.

The implemented latency model only depends on the orbital geometry specifics of the spacecraft involved and hence estimates communications latency from data acquisition to ground reception, regardless of the actual network protocol implementation. Such detachment from a specific protocol is desirable as it makes the analysis more general and the computation times quicker. As explored in the literature, protocols for FSS can be based upon MANET techniques in case of complete absence of coordination and node participation forecast, or in more classic protocols if the network status and topology can be known a priori.

The model for bandwidth allocation starts by computing the total data volume each spacecraft can downlink on a specified time-frame. The spacecraft in the federation add up their downlink volumes as a pool, and then an optimizer re-allocates them to different spacecraft under a minimum bound constraint. The target function of the optimizer is the resulting synergy value. This enables us to compare optimal FoS states against other optimal states. The optimizer implements a derivative-free method based upon Sequential Quadratic Programming (SQP) (Venkataraman, 2009). This bandwidth allocation model is decoupled from the latency estimator and assumes that, in the long term (i.e, several days) it is possible to share an arbitrary volume of data between federates in the network. Table 20 summarizes the assumptions and model parameters that have been used in this evaluation.

Table 20. Models used to evaluate architectures.

Model	**Type/ Reference**	**Parameters**
Orbital propagation	FSS toolkit (Grogan et al., 2014), SGP4	30 seconds step, 5 days simulation
Cost	Models for communications equipment and dedicated stations (Annex II) and real costs for GR and station lease.	See tables
Latency	Geometry based, 500 samples	5 days simulation, 1-hop limitation
Bandwidth	Optimal allocation of global bandwidth pool	5 days simulation, SQP solver, minimum sharing constraint,

The bandwidth allocation in the nominal case assumes a minimum distribution of 10% of the average bandwidth available to federates, per spacecraft. That is, each federate is guaranteed a minimum downlink volume, avoiding cases where the optimal allocation assigns no resources to a particular spacecraft.

6.1.3.2 On models execution time, implementation and validation

The orbital propagator used is based upon Orekit (Maisonobe and Pommier-Maurussane, 2010). It is an established flight dynamics library used by many practitioners. The propagation technique of choice is the proven Simplified General Perturbations-4 model (SGP4), a quick computation model that can capture sun-synchronous orbit specifics. The implementation of the FSS toolkit in Java by Paul Grogan (2014) can run thousands of steps in matter of minutes. To save run time, a scenario with all the potential federates and ground segments is pre-computed beforehand, yielding an adjacency matrix from all assets to all assets at all time steps. During run-time, the adjacency matrix is composed with the ground station assignment and existing federates, to assess the latency of each of the specific architectural instances. The latency estimator performs mostly array manipulation and search operations through the adjacency matrix to establish the occurrence of ground station passes or one-hop relays to other federates. For speed, this specific routine has been compiled in C language.

Typical execution times for the assessment of latency with 500 samples are about 0.02 seconds in machine with Intel Core® i7 4770 @3.4 GHz clock. The latency estimator takes a given amount of samples through the available simulation time. If a sample is taken at each and every time step, and the result averaged, we obtain the exact expectation of latency for data generated constantly during an orbit. To quicken up the evaluation of the architectures we require to reduce the sampling to a few hundreds. Table 21 details the degradation of precision due the sampling.

Table 21. Test case to evaluate the adequate sampling of latency.

Parameter	**Value**	
Satellite	Sat 1 as in Table 14	
Ground Segment	Svalbard and Troll stations	
Propagator Settings	5 day simulation, 30 second time steps (14400 total)	
Latency with Complete sampling (14400 samples)	28.9938 minutes	Error %
10000 samples	28.9345	0.20
5000 samples	28.9439	0.17
2500 samples	29.2338	0.83
1000 samples	28.7795	0.74
500 samples	28.7630	0.80
250 samples	28.8120	0.63
100 samples	29.7600	2.64
50 samples	29.8776	3.05

All samplings perform remarkably well in this example, with larger divergences occurring below 250 samples. A sampling of 500 samples has been selected.

6.1.3.3 Standalone Pareto Front

The models introduced support the evaluation of the architectures enumerated as per Table 19. 57 of the 75 feasible architectures do not feature FSS interfaces and hence are evaluated to establish the standalone system Pareto front. Figure 26 and Figure 27 depict the tradespaces of Utility and Cost for Bandwidth and Latency for the first system. Note that the tradespace is actually three-dimensional, and the Pareto front is a surface as shown Figure 29.

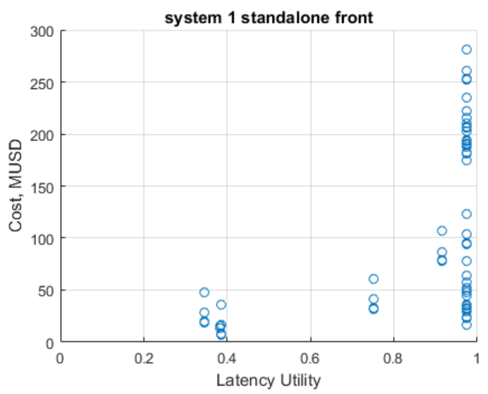

Figure 26. Latency Utility against cost. Note that only 6 standalone options for latency are possible (the 5 GS options and GR).

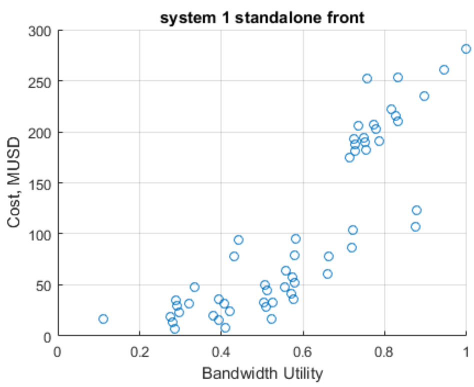

Figure 27 Cost vs. bandwidth utility of System 1.

For brevity we include here only to the tradespace for satellite 1. Due the usage of same utility measures, system architectural enumeration and models, the tradespaces are similar in this nominal case, featuring slightly different results only due the different orbital parameters of the spacecraft included. In the nominal case assessed here, between 19 and 25 architectures are Pareto-Optimal depending on the system under consideration. A surface can be adjusted to these points with different techniques. Figure 28 depicts the surface obtained using linear interpolation and Delaunay triangulation, and Figure 29 shows an adjustment using closest neighbor. Outside range design points are evaluated to the closest neighbor, in both cases.

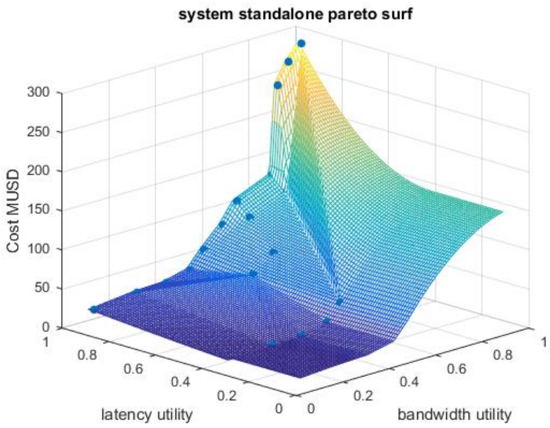

Figure 28. Surface adjustment using linear interpolation and Delaunay triangulation on standalone Pareto-Optimal architectures, an estimator of the Pareto front.

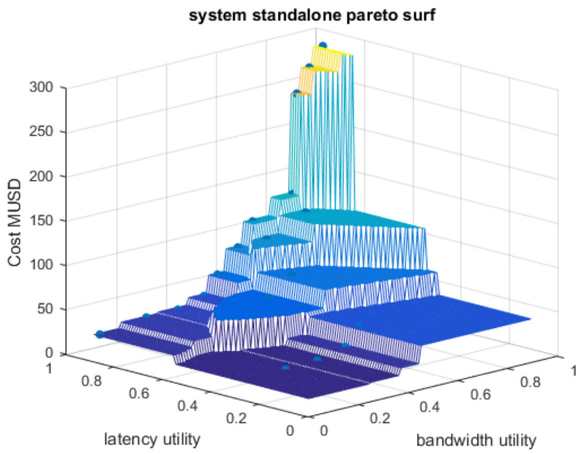

Figure 29. Surface adjustment using nearest-neighbor on standalone Pareto-Optimal architectures.

Choosing between these two interpolation methods to generate the Pareto front surface comes with specific drawbacks. In the case of nearest neighbor, large cost plateaus are created. In these regions it is possible to increase both utilities at constant cost level of equivalent cost. This circumstance is mitigated with linear interpolation, which only features this effect in a few local points. On the following, the linear interpolation technique to generate the Pareto front surface is adopted, as it has a better behaved response to the changes in utilities.

For each system we proceed to generate such surface, which corresponds to the function F defined in Chapter 5 (Eq.5-8). With this, we have finished the preparatory step of the framework, which yields a characterized set of systems, ordered by commissioning date.

6.1.4 Evaluate FoS configurations: compute synergy

This step is the beginning of the exploration phase. It takes the next system to be commissioned as the date advances, and runs an evaluation of its A architectural alternatives, in the FoS context. For each alternative, the cost difference NBC_i^j to the SPF is computed, where i references the system and j the architecture. If there are other systems in the FoS, their NBC_i^j is also updated taking into account the dynamic change in utility due to cooperation, as the framework steps prescribe. Then the overall value of *Sy* is computed as per Eq.5-14. This yields a FoS configuration derived from alternative *j* with a S_y value. The FoS configuration includes the current system's architecture evaluated and the architectures selected for previous systems.

Each subsequent system added adds a set of successor states to each possible previous state. In this manner, the MDP tree, or space of states, is generated. In the case we are exploring, every new system would add 75 sibling nodes, corresponding to all the architecting options. However, all the non-federated architectures collapse to single option as they all present 0 synergy and no couplings with future states. That is, a system has the option to opt-out of federation when being architected, and its actual design becomes irrelevant to the other federates. While all this options were needed to generate the standalone tradespace and Pareto surface, they do not play a role in the exploration of the state space. The MDP graph in this problem is acyclic, hence it is a tree of options.

The case considered presents 19 architectural options which include federation, and the additional non-federated option, for a total of 20 potential successor states. Hence, the full tree in the case we are demonstrating consists of 3,368,420 states, if we recall Eq. 5-25.

$$M = \sum_i^N A^i \qquad \text{Eq. 6-1 (recalls 5-25)}$$

In the last stage only, with five systems deployed, the combinatorial of final FoS configurations is $20^5 = 3.2 \cdot 10^6$. The evaluation of the synergy state in such large amount of FoS configurations is a time consuming process. Besides speeding up each of the state value computations, an option to address this problem is to explore only a part of the MDP tree. One can define a heuristic, or a *policy* in MDP nomenclature, to select nodes and discard others, and hence direct the tree exploration. However this does not guarantee the finding of the best possible path in the MDP. In order to guarantee global optimality, we need to explore larger sections of the tree, if not all the states. The latter is only possible with a reduced computational time when using pruning mechanisms as first hinted in Chapter 5. Section 1.5 compares different policies with full tree exploration. The nominal case introduced here implements a greedy-3 policy as detailed next.

6.1.5 Apply policy to select FoS configurations

After each system's architectural options are explored and the next states in the MDP tree evaluated, we apply a policy to discard less promising successor states in the tree.

We assume system architects of each system favor alternatives that lead to highest distributed and discounted synergy values (Eqs. 5-19, 5-20). This translates into the highest cost advantages in their local tradespace respects to the SPF. We need now also to include in the picture a discount factor, which naturally captures architect preferences for early rewards. In the case of FSS, when systems are commissioned at an approximately yearly cadence and have a design life of about 7-15 years, the discount factor is necessarily to realistically represent architect's decisions. We also consider a cost constraint C_c to restrict the options of each architect, representing preferences for different sections of the SPF, and run sensitivity assessments on this variable. The expression in Eq. 6-2 represents this policy.

$$S = \{s_1, s_1, \dots s_M\}$$

$$S' = \{\, s_k \in \; S \;, \forall k \in \{1, M\}, Sy_k > T, C_{sys} < Cc\} \qquad \text{Eq. 6-2}$$

In Eq. 6-2, S is the set of all successor states s_k being analyzed, their values of synergy Sy_k, and S' is the set resulting from applying a policy on S. C_c is the cost constraint to apply to the architecture of the system being commissioned, which has a cost C_{sys} associated. T is a threshold value of synergy, that in this particular demonstration case we take as equal to the Q-highest value of synergy of the successor states in S. More specifically, in this example we will take $Q=3$.

This type of heuristic is often called *greedy* in tree search assignments as it only takes into account the value of the immediately successor state and does not look further. We call here a $Q=3$ policy greedy-3.

In summary, at each step we select the three FoS configurations with highest synergy, we explore the architecture alternatives of the next system on top of those, and again select the three highest synergy states. This represents only a heuristic and does not guarantee to obtain the highest distributed and discounted synergy, which might be hidden after apparently low-synergy intermediate states.

After all systems have been commissioned and the partial MDP tree built with the corresponding policy, we compute the distributed-discounted synergy of each path based on synergies at each state, as defined in Eq.5-20. Based on this value we can choose a single, winning path for the architectural evolution. The architect of the first system can then chose the architecture alternative that starts this path.

In many cases, before the second system is commissioned, such alternative will be suboptimal in the first system tradespace. Suboptimality is expected; it occurs when interfaces are being implemented before any other federate is present. Suboptimality is measured as a negative NBC of the first system architecture. Posterior systems shall, assuming optimal decisions, include FoS interfaces and henceforth switch the NBC to positive. This introduces a risk from the perspective of the first system architect. This risk of suboptimality is bounded by the cost of the interfaces as discussed in the results section.

6.1.6 Sensitivity analysis for the nominal case

This step includes the exploration of the FoS configurations decision tree with different assumptions and policies. In the example implemented here, we explore the sensitivity to the cost constraint *Cc* in the policy, and to the lifecycle cost of the FSS interface, a key question to understand under which conditions FSS is advantageous for the parties involved. Table 22 lists the parameters assessed.

Table 22. Parameter sweeps for sensitivity analysis in the nominal case.

Parameter	Sweep
FSS interfaces lifecycle cost	[1,2,4,6,8,10,12,14,16,18,20,22,24] MUSD
Cost constraint	[20,40,80,∞] MUSD

6.2 Results of the nominal case

The approach and its implementation in the demonstration case leads to the identification of a winning path for each set of FSS cost and architectural cost constraint parameters. The results are presented as the ratio of discounted and distributed synergy by FSS lifecycle costs, and swept against the latter as described in Table 22. This measure is a Return on Investment (RoI). Figure 30 shows the results of winning architectural paths without any cost constraint limit.

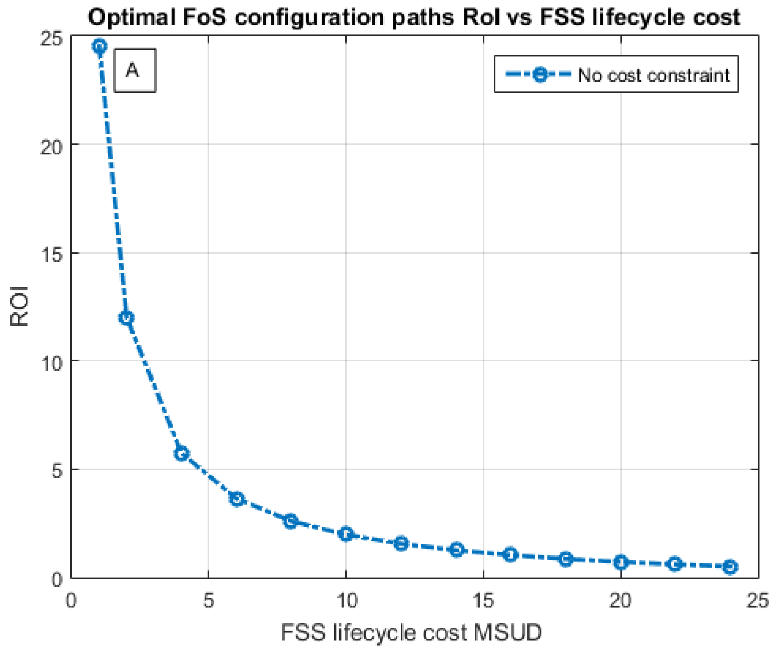

Figure 30. RoI measured as Discounted and Distributed synergy divided by FSS lifecycle cost, from the perspective of first system. Each point represents the winning FoS configuration path under the corresponding assumptions.

The first point on the left in Figure 30 corresponds to a 24.5 MUSD discounted and distributed synergy, obtained when the FSS interface costs only 1 MUSD per 10 year lifecycle. This is the cost advantage experienced by the system through lifetime, in FY16 dollars, taking into account the cadence and the successive commissioning of the other systems. In absence of cost constraints, all of the winning FoS configuration paths are based upon the most extensive ground stations strategy and a 500 Mbit/s link. Point A has been selected to showcase its path (Table 23). Most notably, even a 24 MUSD cost assumption for the FSS cost yields positive synergy, and systems adopt a federation. This stems from the lack of cost constraints that allows system architects to choose high utility-high cost architectures, on the 100-300 MUSD range (see Figure 28). In such architectures the cost footprint of FSS is small. Moreover, due the steeper cost-utility slope in this part of the SPF, the utility increase achieved by federation represents higher cost savings in this area. This is a general insight for systems with this SPF topology, in which cost increases in the high cost region only obtain minimal utility returns.

We explore now the winning FoS configuration paths again, this time with 3 cost constraints at 20, 40 and 80 MUSD. Figure 31 shows the achievable RoIs. In some cases, RoI 0 is reached. This represents a scenario where the highest synergy paths do not actually include federates, no system has interfaces in place, and the highest synergy attainable is 0. When systems adopt architectures from the SPF without federation interfaces, all their cost advantages are 0 and so is synergy. Federating under those conditions would actually yield negative synergy, and hence does not appear in the winning configuration path.

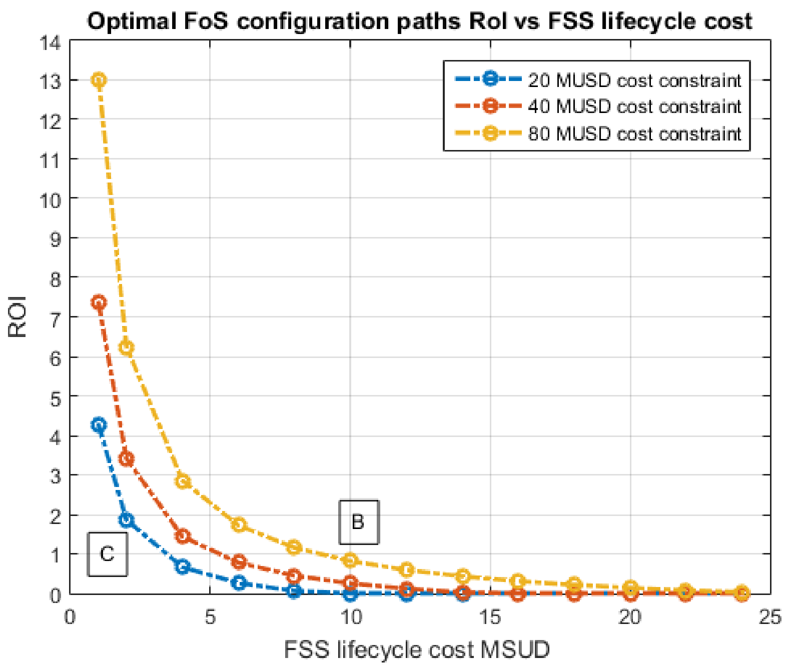

Figure 31. Return on investment measured as Discounted and Distributed synergy divided by FSS lifecycle cost, from perspective of first system. Results for 20, 40 and 80 MUSD architectural cost constraint.

From about 10 MUSD FSS costs attainable synergies are little, and from 15 MUSD onwards federating is not cost-effective, yielding configuration paths without federates as the optimal ones. If FSS interfaces are below 10 MUSD, the attainable RoIs are between 0.5 and 13 times the FSS investment, in this case.

The risks incurred by the first system of the set are bounded by the cost of the federated interface. In case of failure to establish a federation with other missions, this expenditure would not yield any benefit. This is an additional rationale to keep FSS lifecycle costs low. FoS configuration paths B and C, from the 80 and 20 MUSD constraint respectively, have been highlighted and are analyzed in Table 24 and Table 25.

6.2.1 Architectural paths of interest

We examine more in detail FoS configuration paths A,B,C, as indicated in Figure 30 and Figure 31. The first configuration, A, is common across winning paths in Figure 30.

Table 23. FoS configuration path A (No cost limit, 1MUSD FSS).

	S-G Rate Mbit/s	GS architecture	GR	FSS	Cost (MUSD)	Final conf. Local NBC (MUSD)
Sys 1 arch	300	5	No	Yes	88	181
Sys 2 arch	500	5	No	Yes	108	-8
Sys 3 arch	500	5	No	Yes	106	110
Sys 4 arch	500	5	No	Yes	106	-75
Sys 5 arch	500	5	No	Yes	104	172
					Discounted, integrated and Distributed Synergy Value (MUSD)	**24.5**
					RoI	**24.5**
					Synergy value across the 5 states of path (MUSD)	**{-1,62,63,63.5,75}**

FoS configuration path A is a combination of very high cost and utility systems, all of them federated and cooperating. On the final state with the six systems deployed, a synergy of 75 MUSD is achieved, combining all cost advantages present. Note that the first system experiences a change in its utility leading to a 181 MUSD positive NBC. This, added to its actual cost of 88 MUSD, reaches the 269 MUSD mark, close to the limit of the SPF considered in Figure 28. That is, by cooperating, this system achieves nearly the maximum utility levels possible, which would have cost in a standalone fashion an investment of 269 MUSD. All subsequent systems except the last adopt a similar architecture with 500 Mbit/s data rate. The slightly different levels of cost for the latter are caused by variations in the orbital parameters. This drives the amount of passes and, in ground segment strategy 5, costs are based upon the actual satellite passes on the NASA NEN network.

It is noteworthy the fact that system 4 experiences significant disadvantages, before incentives are exchanged. This stems from the orbital constraints and specific visibility patterns of this example as can be seen in Figure 32. System 4 has limited access to federates 1,2, and 3. Hence, it practically establishes an isolated federation with the following system 5, to whom supplies large bandwidth capabilities. Notwithstanding geometric visibility issues with other federates, is still optimal for both to adopt the FSS interface, since their combined cost deltas are larger than 0. This is a case of weak synergy as discussed in chapter 3.

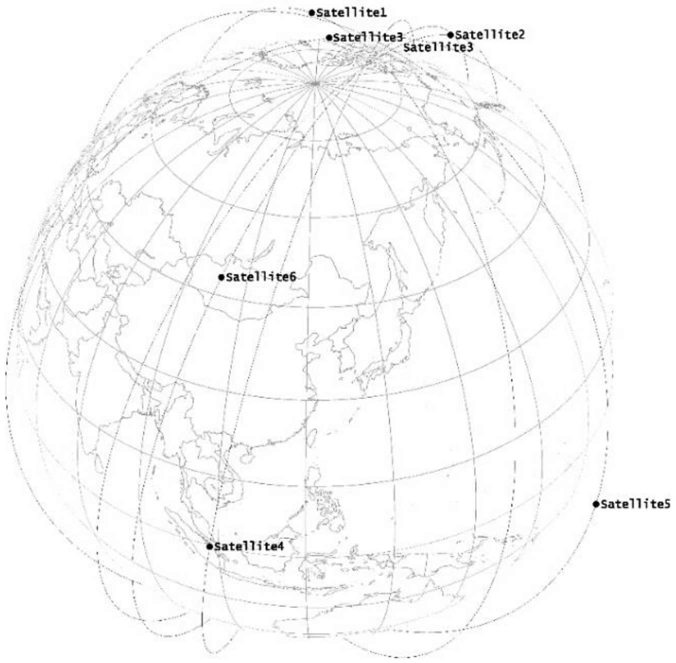

Figure 32. Orbital geometry of the satellite systems included in the nominal case. Satellites 4,5, and 6 are on their ascending pass.

Finally, note that already from the second state, when only Systems 1 and 2 are commissioned with the architectures shown in Table 23, the synergy is positive and all successor states have positive synergy, as shown by the synergy value across states. Next, Table 24 shows the architecture for component system in the FoS configuration path B.

Table 24. FoS configuration path B (80MUSD cost limit, 10MUSD FSS).

	S-G Rate Mbit/s	GS architecture	GR	FSS	Cost (MUSD)	Final conf. Local NBC (MUSD)
Sys 1 arch	500	2	No	Yes	71	-51
Sys 2 arch	500	2	No	Yes	71	-52
Sys 3 arch	300	4	No	Yes	26	155
Sys 4 arch	500	4	No	Yes	46	129
Sys 5 arch	500	4	No	Yes	46	-29
	Discounted, integrated and Distributed Synergy Value (MUSD)					**8.2**
					RoI	**0.82**
	Synergy value across the 5 states of path (MUSD)					**{-10,24.8,25.8,23.8,30.2}**

Path B is remarkable as it features some heterogeneity in the options selected by each architect. The best option for systems 1 and 2 is GS option 2, featuring an arctic and antarctic station. System 3 follows then with GS strategy 4, which is the less expensive option and features less stations in high latitudes. Combining strategy 2 and 4, however, yields a geographically diversified pool of ground stations. Most notably, in this FoS configuration path, when only systems 1 and 2 are deployed they experience positive NBCs, a condition of strong synergy. This condition changes with the addition of the system 3, which chooses a more modest configuration for only 26 MUSD costs but, in virtue of the optimal allocation of resources, obtains an NBC of 155 MUSD. This leads to highest synergy as the third mission's communications architecture gets about 6 times the value it would have in standalone more. This represents a case when system 1 and 2 provide large amounts of bandwidth and contact opportunities to system 3. In this case, mission 3 would need to allocate incentives to the other systems with a minimal amount of 51 and 52 MUSD through lifetime.

System 4 and 5 for the reasons mentioned before do experience lesser values of NBC, but do raise the collective synergy value and would adopt federation nevertheless. This time system 5 acts as supplier of bandwidth and ground access opportunities for system 4.

For the sake of completeness, we now illustrate the RoI experienced not only by the first system, but for all 5 in this case. In order to do that, we project the last stage of synergy up to 5 years ahead –considering no additional federates–. Figure 33 shows how systems 2,3,4 and 5 experience much larger RoIs than system 1 as they do not incur in initial losses. Also note the advantages of joining the federation at more mature stages comparing system's 2-5 RoIs.

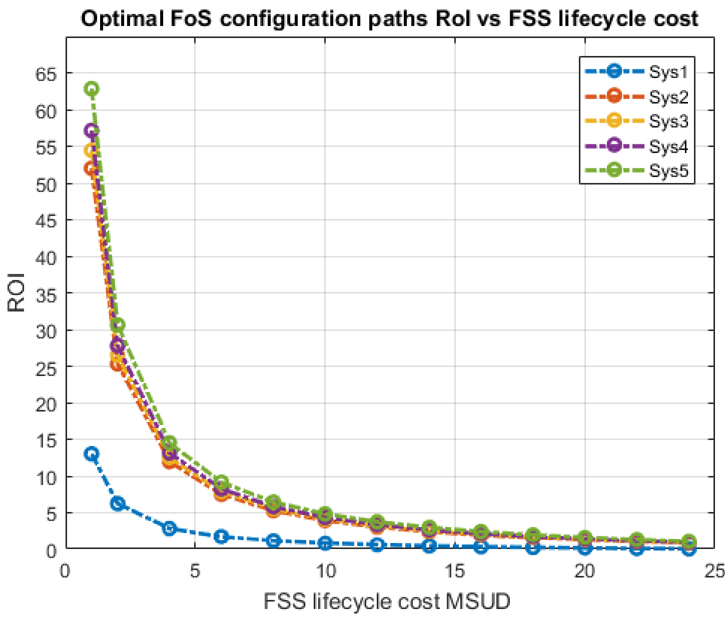

Figure 33. RoI for the 5 systems involved, 80MUSD limit. For ease of comparison we projected the federation up to 5 years after the last system's commission, adding the discounted synergy benefits of the last stage to each RoI accordingly.

Table 25. FoS configuration path C (20 MUSD cost limit, 2 MUSD FSS).

	S-G Rate Mbit/s	GS architecture	GR	FSS	Cost (MUSD)	Final conf. Local NBC (MUSD)
Sys 1 arch	150	4	1	2	10	11
Sys 2 arch	150	4	1	2	10	19
Sys 3 arch	150	4	1	2	10	19
Sys 4 arch	150	4	1	2	10	15
Sys 5 arch	150	4	1	2	10	16
Discounted, integrated and Distributed Synergy Value (MUSD)						**3.7**
RoI						**1.86**
Synergy value across the 5 states of path (MUSD)						**{-2,6,10,14,15}**

Finally, Table 25 shows the FoS configuration path for case C. Case C shows an optimal architecture for all systems of 150 Mbit/s, lease ground station strategy and FSS interface. Most interestingly, this case features strong synergy: due the relatively inexpensive architectures chosen and the low cost of FSS interface, the moderate advantages of federating overcome the costs for all federates, even before incentives exchange. In contrast with case A and B, when some systems deploy expensive ground segments to serve others, in case C there is little additional capacity being deployed and hence no system does incur in losses before incentives. This is a configuration with strong synergy and would not require of incentives exchange.

6.2.2 Evolution of local tradespace of system 1

Now we focus on the first system architect. The results of the previous section recommend various architectures for System 1 to maximize the expectation on discounted returns, depending on FSS cost and cost constraints. For the showcased architectural path A, Figure 34 shows the evolution in the utility-cost space of system 1 through successive commissioning of the other systems and FoS configurations.

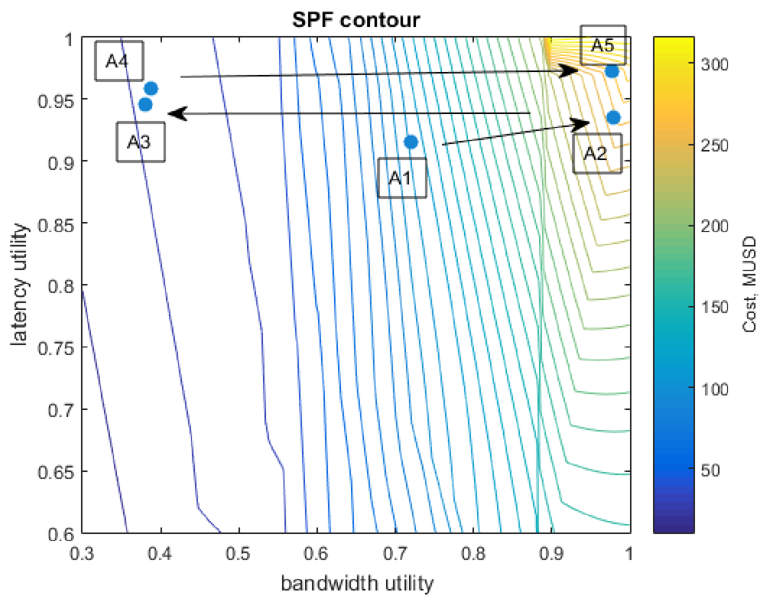

Figure 34. Detail of high utility section of SPF of system 1, and utility changes for system 1 through architectural path A. The chosen contour visualization is equivalent to the 3-dimensional view on Figure 28.

System 1 costs are fixed as the architecture is chosen, and amount to 115 MUSD. After system commissioning and before any federates, system 1 slightly is suboptimal in cost due its investment in the FSS interface, as the colorcode shows in Figure 34, point A1.

After other federates are deployed as per path A, System 1 design point transits to other of utility coordinates, in sequence A2-A3-A4-A5. A2 and A5 dominate the standalone Pareto Front, or SPF, represented by a surface in this case. While latency utility is always increased due to the increase in relay options through the federation, the re-allocations of bandwidth capabilities for optimal synergy can increase or decrease the utility for this particular system. Note how in the sequence A1-A2 and A3-A4-A5 the system is improving in both utilities, while in A3-A4 the system accommodates a utility trade to support the increased amount of federates. Notably, with the addition of the 3rd federate, system 1 releases significant bandwidth to support the former. This is recovered as the last system comes in to support. Again, note the fact that while A3 and A4 configurations are locally and temporarily suboptimal for system 1, this does not mean system 1 would opt out of the federation, because through all the sequence A2-A5 synergy is positive (see table Table 23) and this by definition means such local suboptimality is outweighed somewhere else in the federation and the overall net effect is positive. That is, the generated value would justify to incentivize this system above its cost losses. Moreover, the entire sequence brings a lifetime RoI for the first federate of 24 times the investment.

In the case of architectural path B, Figure 35 shows how the latency utility increases with at a steeper rate than on the local SPF, but the system is actually offering bandwidth capabilities to other federates. Compared to the SPF, the system is experiencing suboptimal utility coordinates –the architecture is too expensive for the level of performances achieved– but as the overall synergy is always positive, such cost overrun would be compensated by other federates. As discussed before, in this case, system 3 has a margin of more than 150 MUSD to allocate incentives.

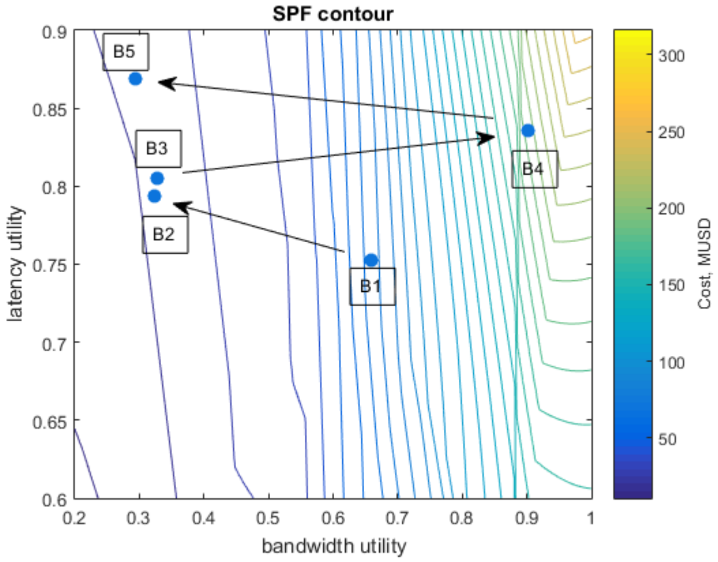

Figure 35. Detail of of SPF of system 1, and utility changes for system 1 through architectural path B.

Finally, the same representation for path C, at 2 MUSD cost, yields a similar profile. Figure 36 shows the change of utilities. In this case the latency also monotonously improves with the amount of federates, and the bandwidth allocated to system 1 varies slightly from one state to the other. However in all steps, the performances for system 1 are better than what is achievable in local SPF. As the colorcode shows, the architecture points C2 to C5 present a lower cost than what is achievable in the SPF contour.

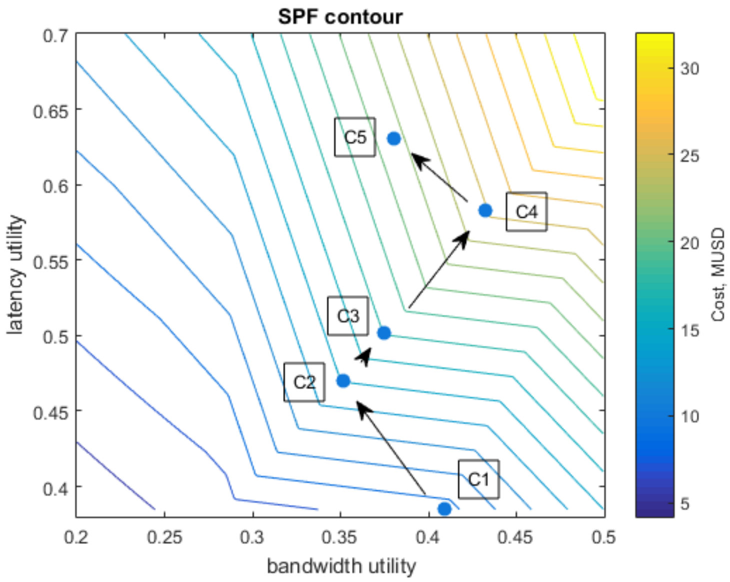

Figure 36. Detail of medium utility section of SPF of system 1, and utility changes for system 1 through architectural path C. Configurations C2 to C5 are always above the SPF.

6.2.3 Conclusions for the nominal case

These results demonstrate the method proposed herewith for the case of a 5-satellite set, showing attainable synergy from the second mission launched and the effects of network topology in potential cooperation exchanges. The specifics of this case suggest that FSS interfaces, including ISL payloads, OBDH, and the associated power and attitude control subsystem changes in the spacecraft shall be below 10 MUSD to enable federations. The results show that federating to exchange bandwidth capabilities and offer relay options is more cost-effective than deploying independent ground segments or using geostationary relay options. Indeed, agency mission portfolios do already share ground segment capabilities (Harris et al., 2016) albeit yet not through ISLs.

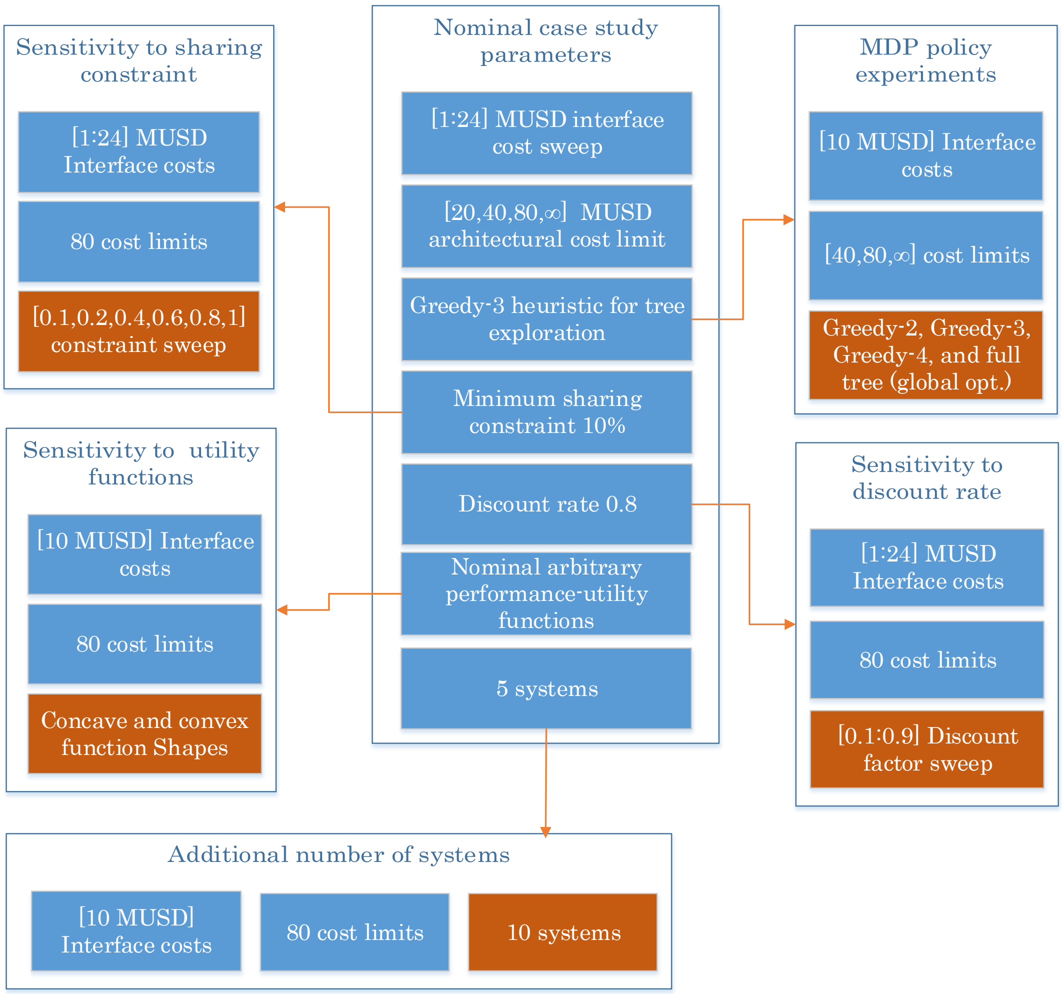

Figure 37. Case study map depicting the sensitivities and additional experiments derived from the notional case.

This nominal case also delivers general insights as it exemplifies how federating can assist systems in achieving larger utilities beyond what is possible on their local tradespace, on a realistic case. The main showstopper for the achievement of cooperation benefit are naturally the interface costs as discussed above, and the network topology, which is only a challenging issue with a reduced amount of systems, and particularly in space scenarios. The framework and models introduced here allow taking into account both issues and quantify the cost advantages of federating, and optimal architectural paths. Now we derive several additional assessments from this nominal case. Figure 37 shows the additional studies derived from the nominal case study.

The next section details the experiments performed with the MDP policy. Sensitivity assessments for discount rates, utility functions, sharing constraints and number of systems follow.

6.3 MDP Policy experiments

First, this section compares several greedy policies, specifically greedy-1, greedy-2, greedy-3 and greedy-4 with the optimal full tree exploration, which has been attempted in two occasions.

6.3.1 Greedy approach

Figure 38 exemplifies the application of the greedy-2 heuristic.

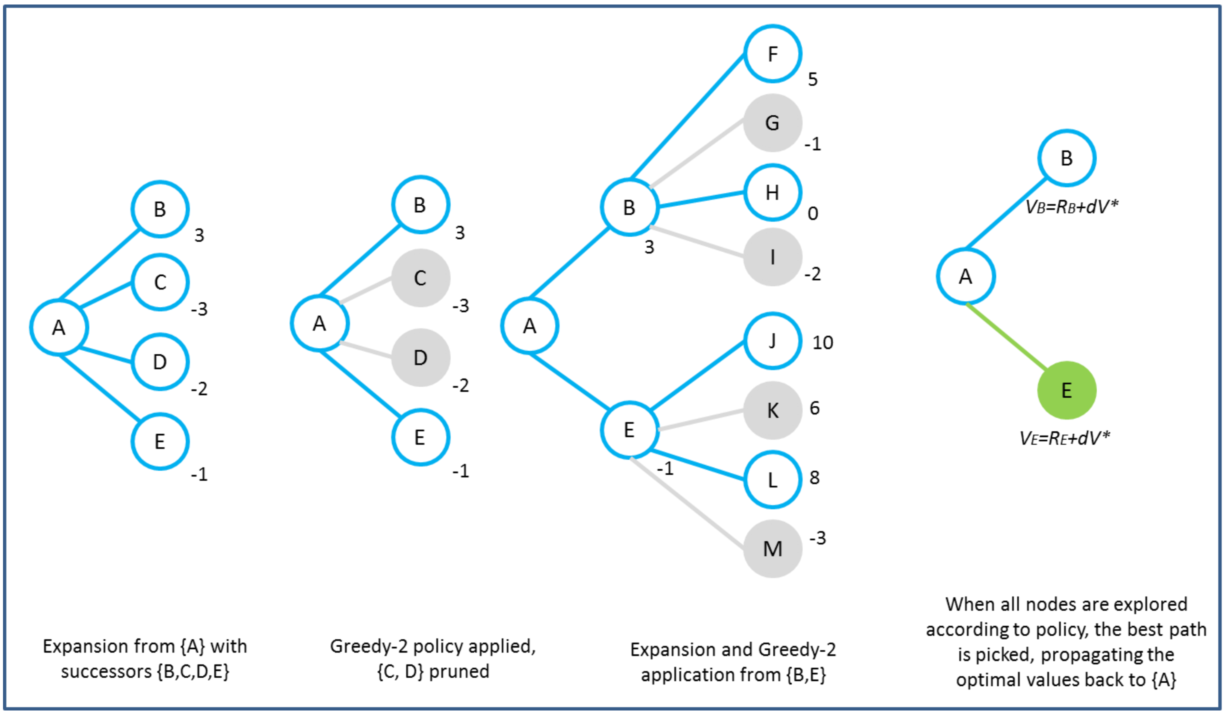

Figure 38. Example of greedy-2 algorithm to partially explore the MDP.

Strictly speaking in MDP terminology, when choosing more than one successor state (2,3,4…), the greedy heuristic is not a single policy but rather a family. As a policy dictates what unique state to adopt next, the exploration proposed here is rather a test on several policies that could be narrated as as *"pick always the best of successor states", "pick the best of successor state on the first choice, and the second best in all successive choices*", etcetera. However for simplicity and clarity purposes we refer to this heuristic as an MDP policy in singular. By using this heuristic, we build only a promising part of the MDP and not the full option tree. greedy-2 creates 2^N paths, where N is the number of tree levels or equivalently, the amount systems commissioned. As an exception, in the first level of the path, where all rewards of deploying FSS are negative and identical, we need to explore all the states without any pruning. Hence, the number of explored states is $N \cdot 2^{(N-1)}$.

After we created all these paths with the greedy procedure, we then proceed to pick the best one. This is achieved by starting from the best successor at the final level, and working our way backwards. Remember the value of a state depends on the rewards obtained in that particular state and the discounted values of the successor states with maximum value *V**. The last step in Figure 38 notionally represents the evaluation of B and E in order to pick a path from state A. If we assume a discount factor γ of 0.8, the value of these states is:

$$V_B = R_B + \gamma V^* = 3 + 0.8 \cdot \max(5,0) = 5 \quad \text{Eq. 6-3}$$

$$V_E = R_E + \gamma V^* = -1 + 0.8 \cdot max(10,8) = 7 \quad \text{Eq. 6-4}$$

Hence in this notional example described with Eq. 6-3 and Eq. 6-4 , *A* shall pick the actions leading to state *E*.

6.3.2 Full tree exploration for global optimality

An alternative way to explore the tree is to compute it entirely and apply the procedure to find the best path on all the options, instead of the subset rendered by a greedy policy. With the aim of speeding up this process and computing as little state values as possible, we use three types of pruning techniques when exploring tree mechanism: forwards, lateral and backwards. In order to prune a branch of the tree without loss of a potential global optimal state hidden downstream, we must use best and worst-case bounds for pruning, in a manner resembling the alpha-beta pruning technique (Russell and Norvig, 2002). Call the maximum state value achievable *Vb*, the lowest possible value a state can have *Vl*, and the actual, final value of a state computed after exploration of all downstream possibilities, *Vf*.

Hence we can *forward* prune sibling states by comparing the *Vb* and *Vl*, amongst them, before any of their successors have been computed. If a state best prospect *Vb* is inferior to any other state's worst prospect *Vl*, this state can be safely removed from the exploration queue. Formally the set S' states can be pruned from the set of siblings' S if:

$$S = \{s_1, s_2, \ldots s_M\}$$

$$S' \subseteq S, S' = \{s'_1, s'_2, \ldots s'_K\} \quad \text{Eq. 6-5}$$

$$\forall s'_i,\ \exists s\ \in S, Vb_{s'_i} < \ Vl_s$$

The value of *Vb* in Eq. 6-5 can be computed using the discount factor, and the expected maximum synergies derived from the case in question. *Vb* is a geometric series featuring the addition of discounted upper bound future rewards, limited by the lifetime of the system. That is:

$$Vb = \sum_{i=1}^{lc} \frac{r_{max}}{lc} \cdot \gamma^{(i)} = \frac{r_{max}}{lc} \frac{1-\gamma^{lc+1}}{1-\gamma} - \frac{r_{max}}{lc} \qquad \text{Eq. 6-6}$$

$$Vl = 0 \qquad \text{Eq. 6-7}$$

In Eq. 6-6, *N* is the total number of systems deployed, r_{max} is the maximum reward attainable per system, γ is the discount factor. r_{max} depends on the cost limits and minim architectural costs. That is, in the case the maximum cost attainable in the SPF is about 300 MUSD, and the minimum cost of an architecture is 6 MUSD, in the best of cases, synergy would be 300-6=294 MUSD. This is the utopian point on which the architecture with the least initial value experiences a shift in utilities to the absolute maximum of the SPF. While this is hardly attainable in practice for a system, let alone for all participant systems simultaneously, it still has the capability to prune some branches of the tree. The expected lifecycle of the system is *lc*. If the deployment cadence of the systems is different than a year, and rewards evolve during the year, r_{max} needs to accommodate such changes; or the series expression needs be rewritten to add terms more than once per year. As Eq. 6-7 notes, *Vl* is 0; that is, in the worst case no future rewards will be attained, and the system will continue to experience the current state value indefinitely. The pruning bounds *Vl* and *Vb* at each step need to be cohesive with the system lifecycle of the system being deployed.

6.3.2.1 Forward pruning

Figure 39 exemplifies this case of *forward pruning*; states B,C, and D have been evaluated and their rewards *R* are known, but their final Value *Vf* is not available, pending exploration of their successor states. However, if the value of the successor states can be bounded, an expectation for best case and worst case can be performed.

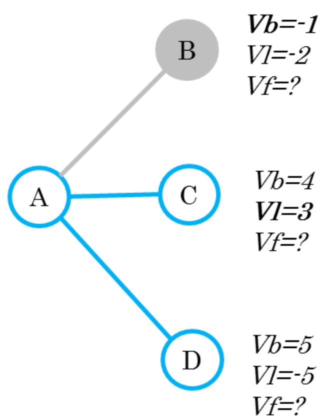

Figure 39. Example of forward pruning

In this example, the best value expectation for node B is -1, less than the worst expectation for C, hence B can be safely pruned. Note how states C and D cannot safely prune each other based on the bounded expectations.The exploration proceeds then depth-first, expanding from one of the surviving states in an attempt to establish its final *Vf* value. If a state is assigned a *Vf*, we re-apply a pruning mechanism before expanding the other sibling states. This mechanism we call *Lateral Pruning*.

6.3.2.2 Lateral pruning

If a *Vf* for a state is known, we can again examine the siblings of the state in the light of this new information. Shall the *Vf* of a state be larger than the *Vb*, the best possible outcome for another, the latter can be safely pruned before proceeding to explore it.

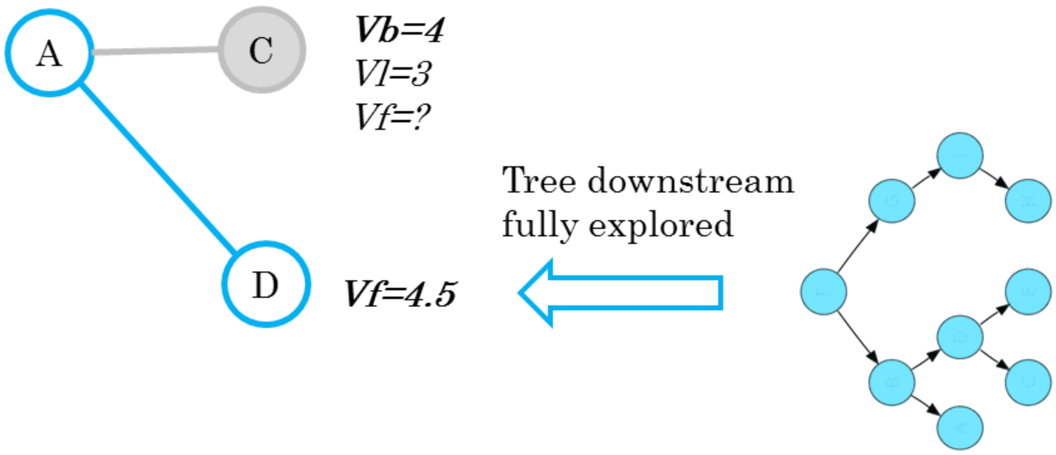

Figure 40. Example of lateral pruning. Upon full downstream exploration, state D can be assigned a final value corresponding to its reward and discounted value of the best downstream option. This prunes C.

The downstream of state C as depicted in Figure 40 had not been explored yet, as we are following a depth-first approach, as corresponds to alpha-beta pruning. After the exploration of the successors of D is finished and *Vf* assigned, we can prune of C as its best expectation is lower than D's actual value. In order to enhance the possibilities of lateral pruning happening, the most promising states (highest current reward) are explored first.

6.3.2.3 Backwards pruning

Finally, as in the case of the greedy approach, when all sibling states have been fully explored, the only remaining action is to choose the best and prune all others based on largest Vf. Then this value is discounted and propagated to the parent node, and the process repeated. This algorithm returns a unique best path.

```
function EXPAND(parent) returns Q'
    If lastlevel=true
        Parent Vf ← parent reward
        Return
    Else
        For each successor state
            COMPUTE successor reward, Vb, Vl
            Q ← successor
        End
        Q ←  FWD-PRUNE(Q)
        Q ← ORDER-BY-REWARD(Q)
        While Q is not empty
            EXPAND(first element in Q)
            Q ← LATERAL-PRUNE(Q)
            Q' ← first element in Q
            DELETE first element in Q
        End
        Q' ← MAX (Q', Vf)
        Return Q'
End
```

Figure 41. Pseudocode featuring the pruning mechanisms (FWD-PRUNE, LATERAL-PRUNE and MAX) in the tree expansion function.

The pseudocode snippet in Figure 41 summarizes the tree exploration and the pruning mechanisms in place.

6.3.3 Comparison of heuristics

The nominal case study introduced in this chapter has been computed with the greedy-3 policy. Now we compare greedy-2, greedy-3, greedy-4, and full tree exploration for 3 different cases extracted from the nominal case. First, Table 26 compares these different heuristics in the case of a 40MUSD cost limit, 10MUSD FSS costs. For reference we include in the tables here the FoS configuration, in a condensed notation describing the architecture of each system. See section 6.1.3 for the descriptions of the architectural options.

Table 26. Comparison of different heuristics to explore the MDP for a 10 MUSD FSS cost, 40 MUSD architecture cost limit.

	Greedy-2	**Greedy-3**	**Greedy-4**	**Full tree (global opt)**
Sy path across states MUSD	{-10,-1.2,15,17,19.7}	{-10,-1.2,15,17,19.7}	{-10,-1.2,15,17,19.7}	{-10,-1.2,15,17,19.7}
Discounted Sy MUSD	2.507	2.507	2.507	2.507
FoS configuration	all {300,4,1,2}	all {300,4,1,2}	all {300,4,1,2}	all {300,4,1,2}
Total states evaluated	560	1,850	4,410	65,610 (approx. 40% prune efficiency)

All heuristics in this case perform identically and find the global optimum solution as returned by the full tree, which is a FoS configuration of all systems with 300 Mbit/s downlink rate, ground segment strategy 4, no GR and FSS. Discounted synergy amounts 2.5 MUSD. The total amount of nodes evaluated is directly proportional to the computation time required, which will also depend on the machine and implementation used. With the implementation discussed in section 6.1.3.2, an Intel Core® i7 4770 @3.4 GHz machine needs about 0.7 seconds to evaluate 10 states, using its 4 cores in parallel.

Given the cost limit of 40 MUSD, only 10 architectural alternatives per system need to be assessed, making the greedy and the full tree results naturally close. Note that a greedy-1 heuristic is not considered as for the first system, opting out of FSS is always better in a one-step look-ahead that all other alternatives. Note the difference in the amount of states that need to be evaluated before returning the best path for the different heuristics; the full tree approach in this case requires of computing about 100 times more states than greedy-2. Also note the remarkable advantages of the pruning mechanisms implemented, which avoid the assessment of more than 25,000 states (40% in this case) in the full tree exploration. Next, Table 27 compares again the heuristics, on the case B described in Section 6.2.1.

Table 27. Comparison of different heuristics to explore the MDP for a 10 MUSD FSS cost, 80 MUSD architecture cost limit. This corresponds to scenario B as explored in the nominal case.

	Greedy-2	**Greedy-3**	**Greedy-4**	**Full tree (global opt)**
Sy path across states MUSD	{-10,24.8,25.8,23.8,30.2}	{-10,24.8,25.8,**23.8,30.2**}	{-0,24.8,25.8,**22,35**}	{-0,24.8,25.8,22,35}
Discounted Sy MUSD	8.2	8.2	8.4	8.4
FoS configuration	{500,2,1 ,2} {500,2,1,2} {300,4,1,2} {500,4,1,2} {500,4,1,2}	{500,2,1 ,2} {500,2,1,2} {300,4,1,2} **{500,4,1,2}** **{500,4,1,2}**	500,2,1,2} {500,2,1,2} {300,4,1,2} **{500,2,1,2}** **{500,2,1,2}**	{500,2,1 ,2} {500,2,1,2} {300,4,1,2} {500,2,1,2} {500,2,1,2}
Total states evaluated	2896	7,840	17,152	810,000 (approx. 28% prune efficiency)

In this case, 16 alternatives are possible for each system. Greedy-2, and 3 obtain the same results. However, the full tree obtains a better solution, slightly increasing synergy. greedy-4 also captures this configuration path, which is the global optimum in this case. The difference between the first and second pair of heuristics can be located in last two steps of the configuration path, corresponding to adding system 4 and 5. Greedy-2 and 3 choose highest next-step rewards in step 4 (23.8MUSD, marked bold in Table 27). Conversely, the last pair of heuristics accept lower synergy in this step, enabling them to find a path to a higher reward in the last step. After applying the discount, this path still yields larger synergy returns.

Another angle to describe this is that the policies favoring shortsighted decisions tend favour systems' architectures that to 'exploit' the FoS rather than 'expand' its capabilities. The architectures of system 4 and 5 with the greedy-2 and 3 heuristics use ground segment strategy 4, lease, which is less expensive and less capable than strategy 2 (high latitude, dedicated, arctic and antarctic station). Since there is a lot of capabilities already present in the FoS, greedy-2 and 3 settle for reducing local expenses and making use of federate resources. Instead, the second pair of policies still explore the high capability option for system 4, enabling a higher synergy in the last step.

If larger discounts would be applied it is likely that shortsighted policies would converge rapidly to global optimum. This behavior has been noted in the discount rate sensitivity experiments in section 6.4.2. Finally, Table 28 shows the results of greedy-2,3,4 for a case without architectural cost limitation. This generates 20 feasible architectural options offspring of each state. Such vast state space has not been computed with full tree approach.

Table 28. Comparison of different heuristics to explore the MDP without cost limit, 10 MUSD FSS cost.

	Greedy-2	**Greedy-3**	**Greedy-4**
Sy path across states MUSD	{-10,53,8,54.1,54.5,66.8}	{-10,53,8,54.1,54.5,66.8}	{-10,53,8,54.1,54.5,66.8}
Discounted Sy MUSD	19.7	19.7	19.7
FoS configuration	{300,5,1,2}{500,5,1,2}, {500,5,1,2},{500,5,1,2},{500,5,1,2}	{300,5,1,2}{500,5,1,2}, {500,5,1,2},{500,5,1,2},{500,5,1,2}	{300,5,1,2}{500,5,1,2}, {500,5,1,2},{500,5,1,2},{500,5,1,2}
Total states evaluated	5,380	14,420	30,920

For the case, greedy 2,3, and 4 present the same results. In the light of the assessment of the heuristics, as a compromise and for consistency to the results of the nominal case, we adopt greedy-3 for the application of the framework in the following section. In doing so we accept the possibility of missing global optima and slightly underestimating the achievable synergy. However, this lack of precision is on the conservative side of the federation evaluation.

6.4 Additional Sensitivity assessments

This section introduces the additional framework evaluation under different utility function assumptions, discount rates and minimum sharing constraint.

6.4.1 Influence of utility functions

In the following, we change the arbitrary utility functions shown in Figure 24 and Figure 25 from convex and concave to their negatives, i.e, concave for bandwidth perfomance utility mapping and convex for latency to utility.

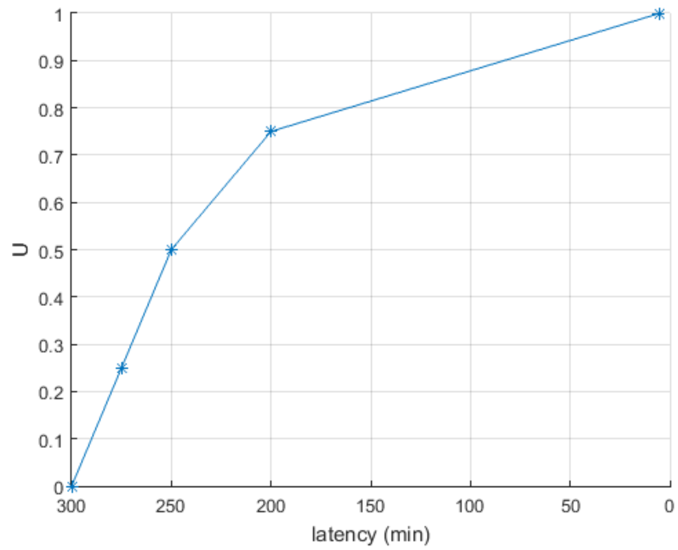

Figure 42. Alternative Utility-latency mapping using a 4-piece concave function, negative of Figure 25.

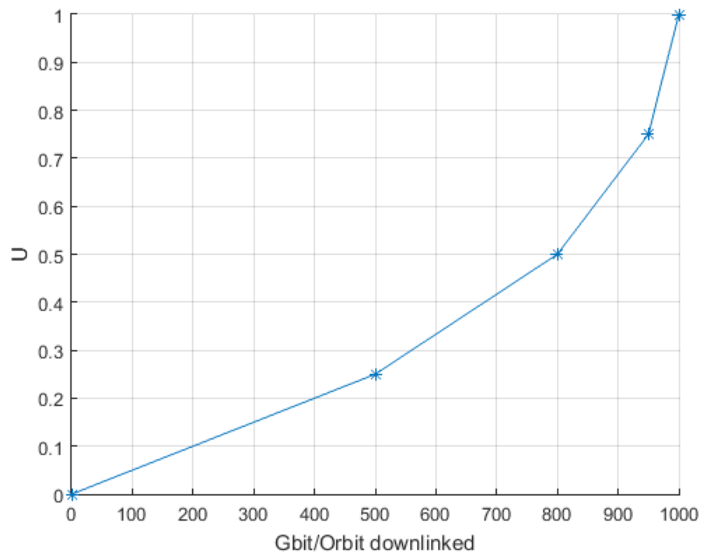

Figure 43. Alternative bandwidth-utility mapping using a 4 piece convex function, negative of Figure 24.

The full framework has been run again, for all FSS cost and architecture cost limit sweeps depicted in the nominal case, with the alternative utility functions. The results, as shown in Figure 44, exhibit little difference with the nominal case. The difference stems from the changes in latency–utility mapping. In the scenarios analyzed, the systems' latency performances mostly reside in the 60 to 0 minutes region, while there is diversity in the bandwidth performances attained. The alternative utility function for latency assign to such performances the range 0.9 -1, in contrast with the nominal, that ranges 0.5 to 1 for such levels of latency. Hence the benefits of improving the systems' latency are more constrained in the former case.

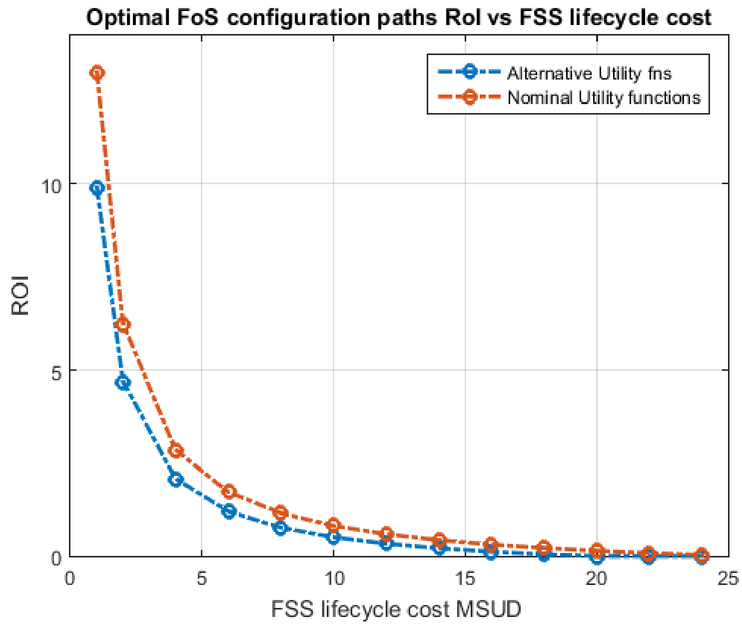

Figure 44. Comparison of the results for 80 MUSD architectural cost limit, 10 MUSD FSS cost, with the nominal utilities and with an alternative utility set as per Figure 42 and Figure 43.

6.4.1.1 Heterogeneous utility functions

Besides changing the utility functions for all systems, we can now use different utilities for the systems involved. The goal of this exercise is to study the emergence of a federation in presence of systems which have different measures of utility, and in this particular case study, to look at the mix between missions in need of state-of-the art performances and missions which do not require large bandwidths or quick data downlink capability. That is, explore federations were some systems have stringent requirements and other present more relaxed need for bandwidth and latency performances.

In order to analyze such case, we now apply to the case B (80 MUSD arch. cost limit, 10MUSD FSS cost) two different sets of utility functions as depicted in Figure 45 to Figure 48. These shapes are named "stringent requirements" and "relaxed requirements".

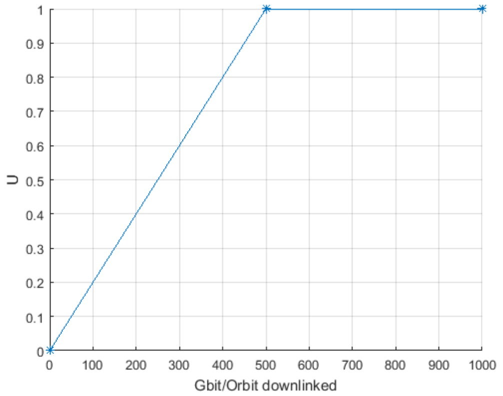

Figure 45. « Relaxed Requirements » utility shape for bandwidth performance. From 500 Gbit/Orbit, utility 1 is reached and the Utility response is flat.

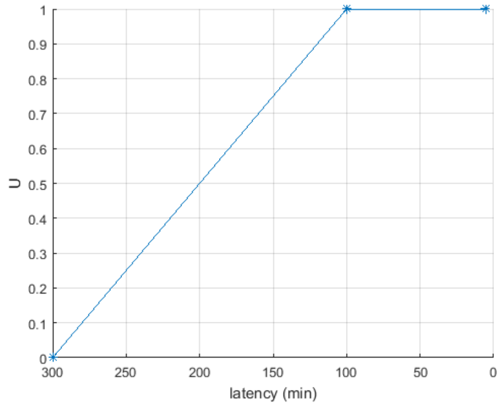

Figure 46. « Relaxed Requirements » utility shape for latency performance. From 100 minutes, utility 1 is reached and the Utility response is flat.

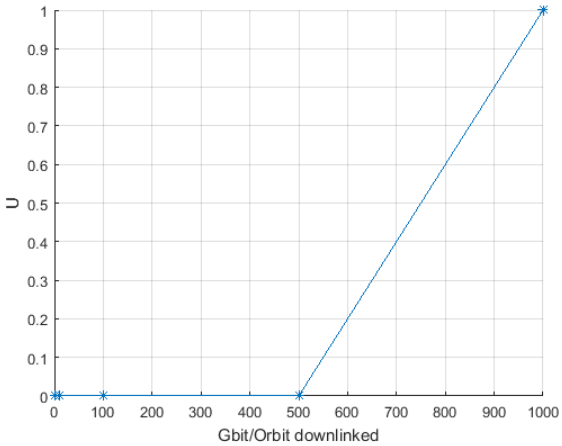

Figure 47. « Stringent Requirements » utility shape for bandwidth performance. Bandwidths below 500 Gbit/Orbit do not yield any utility to the system.

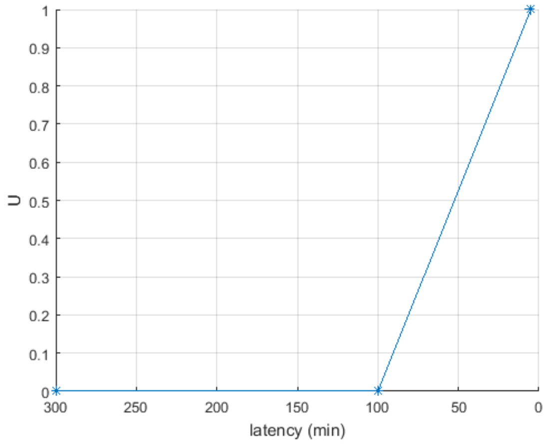

Figure 48. « Stringent Requirements » shape for latency performance. Latencies below 100 minutes do not yield any utility to the system.

Systems 1,3 and 5 will run their architectural exploration based on the Relaxed Requirements approach, while systems 2 and 4 will use the Stringent Requirements shapes. Table 29 summarizes the results and comparison with the nominal case.

Table 29. Scenario with heterogeneous utility functions vs. Nominal case.

	Nominal case	**Heterogeneous utilities**
Discounted, integrated and Distributed Synergy Value (MUSD)	8.2	4.9
RoI	0.82	0.49
Synergy value across the 5 states of path (MUSD)	{-10,24.8,25.8,23.8,30.2}	{-10,10.6,23.8,21.7,14.5}

While the architectural configurations resulting from this exercise are fairly similar to the nominal case, there are differences in RoI and Synergy. In the heterogeneous utilities case, systems 1,3 and 5 do not experience performance benefits from federation: their standalone architectures are enough to achieve the maximum utility levels, given its modest requirements. Instead, said systems support the bid of systems 2 and 4 for large bandwidth and latency. Naturally, when there is less overall response to performance enhancement, the overall benefits of federation are limited. Note, in this case, the values for Synergy are cut in 40%, but 60% of the federates do not really require additional performances and purely act as service providers. Therefore, if its needs are pressing enough and large costs advantages achievable, it is not unthinkable a single system could justify a federation.

6.4.2 Influence of discount factor

The discount factor plays a role in the achievable discounted synergy and the system architects lenience to design their systems for federation. Figure 49 portrays the changes in achievable federation RoI for the first architect as larger discount factors and varying FSS lifecycle costs are applied. The $\gamma=0.8$ discount factor (equivalent to 20% discount rate) corresponds to the previous results, scenario B of the nominal case.

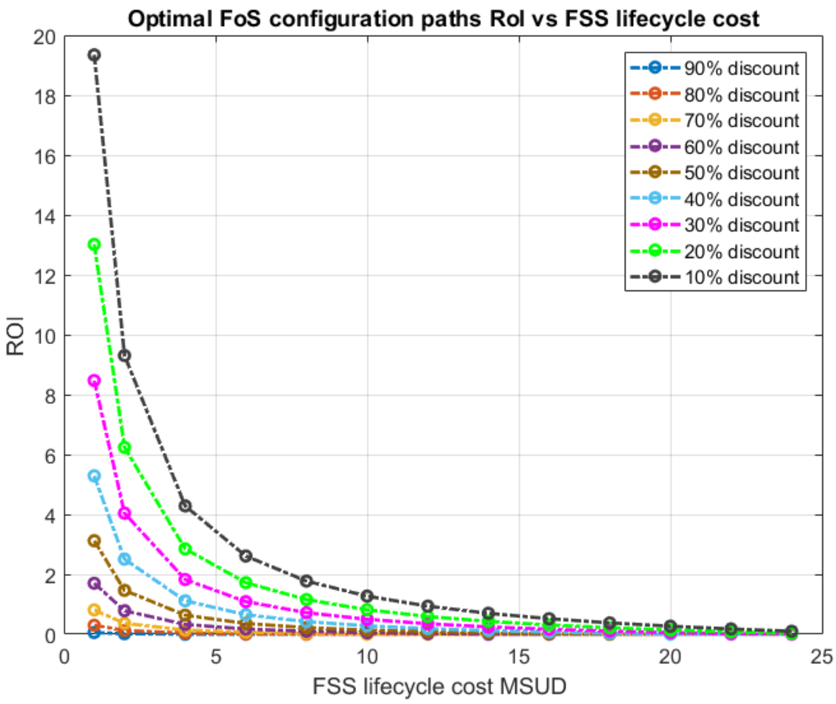

Figure 49. Sensitivity of the achievable RoI of federation to the discount, expressed in discount rate %. Run with 80 MUSD cost limit.

Under the assumptions of scenario B, and for discount rates of 90% to 60%, federating always returns less than twice the investment, and is discouraged from 5 MUSD FSS lifecycle costs. For other discount rates, federating is still appealing up to 10 or 15 MUSD. Note how linear changes in the discount rate do have non-linear returns, an expected behavior of composing discount factors. Shortsighted architects, applying discount rates of 50% or more, expect their system to quickly loose value and hence changes in its performance at late stages of the lifecycle are of little interest to them. Remarkably, all FoS configurations across Figure 49 are the equal or very similar, showcasing same architecture selections as of scenario B. That is, the optimal architecture remains the same across different discount scenarios, while the achieved benefit naturally varies.

6.4.3 Influence of the minimum sharing constraint

The minimum sharing constraint acts as a lower bound of the bandwidth allocation amongst federates, and can potentially restrict the synergy rendered by an FoS. A Minimum Sharing constraint (MS) of 0 allows the bandwidth assignment to completely choke a mission´s output, while a MS of 1 forces all available bandwidth to be allocated equally amongst federates. Figure 50 shows the achievable RoIs again for different MS constraints.

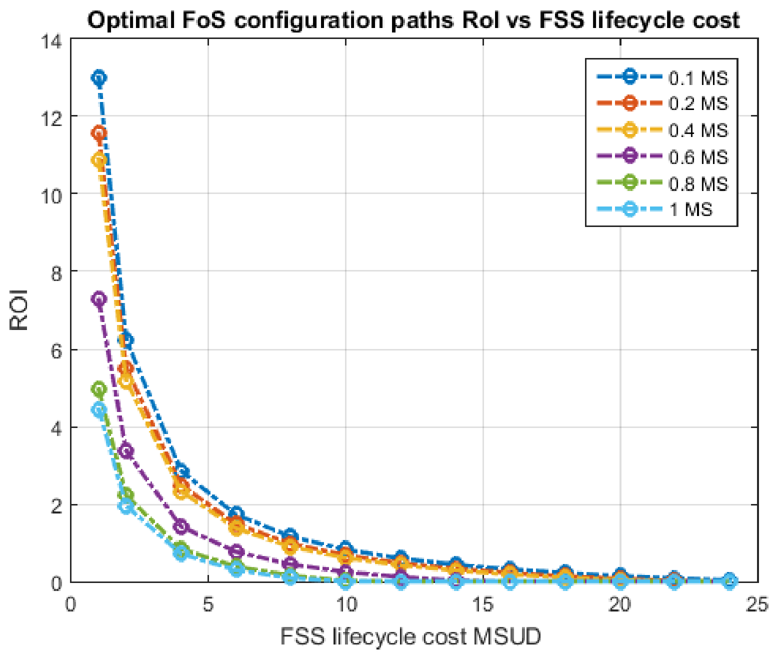

Figure 50. Sensitivity of the RoI results for the 80 MUSD cost limit scenario to minimum sharing constraint MS.

As expected, as we constrain the optimizer, less synergetic solutions can be found. However note the 0.2 and 0.4 MS constraint achieve very similar results. The only breakthrough possible from both is to relax the MS further to 0.1. With this low limit on the minimal bandwidth allocation per federate, the optimization algorithm is able to surpass the RoI 12 threshold. This is achieved by allocation very little bandwidth to some federates and maximum utility bandwidth levels to others.

6.4.4 Additional number of systems

In order to gain generality in the assessments, a case with additional 5 satellites is introduced here. Said satellites are also deployed with a yearly cadence, until 10 years have passed from the commissioning of the first systems. Table 30 details their orbital parameters, also taken from sampling the statistical distribution of LEO orbital parameters of existing missions (Lluch et al., 2015).

Table 30. Orbital parameters and lifetime of additional 5 satellites.

ID	SMA	Ecc	Inc	AP	RAAN	MA_0	BOL (yr)	EOL (yr)
Sat6	7094384	0.000414	98.26	108.16	219.47	224.42	6	16
Sat7	7067740	0.00033	98.15	111.15	235.28	193.40	7	17
Sat8	7175561	0.000386	98.59	117.08	210.17	88.33	8	18
Sat9	7032678	0.00163	98.01	121.77	297.55	143.33	9	19
Sat10	7071689	0.000016	98.16	112.76	38.10	149.73	10	20

The parameters for this framework run are equivalent to the nominal case as depicted in Figure 37; the discount factor is 0.8, the minimum sharing constraint 0.1, and a cost limit of 80 MUSD and FSS cost of 10 MUSD has been chosen for ease of comparison with other cases. The tree exploration implements the greedy-3 policy. Due the additional systems, the execution time for this case is significantly longer; greedy-3 evaluated about 800,000 states before terminating.

An interesting feature to note is that, while the architecture selected for the last system does not directly affect the RoI of system 1, which is at end of life, it does influence it through the other systems. Last system's decisions affects the RoI achieved by the systems in between first and last and hence indirectly does influence the decisions, and RoI of the first. Table 31 shows the results of this case, compared with the shorter 5 satellites nominal run.

Table 31. Winning FoS configuration path for 10 satellites, compared to 5 in parenthesis.

	S-G Rate Mbit/s	GS architecture	GR	FSS	Cost (MUSD)	Final conf. Local ΔC (MUSD)
Sys 1 arch	500 *(500)*	2 *(2)*	No	Yes	71 *(71)*	-48 *(-51)*
Sys 2 arch	500 *(500)*	2 *(2)*	No	Yes	71 *(71)*	-48 *(-52)*
Sys 3 arch	300 *(300)*	4 *(4)*	No	Yes	26 *(26)*	1 *(155)*
Sys 4 arch	300 *(500)*	4 *(4)*	No	Yes	26 *(46)*	181 *(129)*
Sys 5 arch	500 *(500)*	2 *(4)*	No	Yes	71 *(46)*	-51 *(-29)*
Sys 6 arch	500	2	No	Yes	71	198
Sys 7 arch	500	2	No	Yes	71	-48
Sys 8 arch	500	4	No	Yes	46	-28
Sys 9 arch	300	2	No	Yes	51	-28
Sys 10 arch	500	2	No	Yes	71	207
Discounted and Distributed Synergy Value (MUSD)				**6.9 (8.2)**		
RoI				**0.69 (0.82)**		
Synergy value across the 10 states of path (MUSD)				**{-10, 24.8, 25.8, 26.8, 23.8, 32.3, 31.9, 34.8, 34, 33.5}** *{-10,24.8,25.8,23.8,30.2}*		

The architectural choices for systems 4 and 5 vary slightly, but not dramatically, influenced by the future commissioning of systems 6 to 10. The influence of the latter does not seem strong enough to force changes in the decisions taken by the first 4 architects. However, as seen by the first system, there is a slight decrease in RoI. This is caused by the redistribution of resources in late stages of the federation and the architectural decisions of systems 4 and 5. Note how Synergy value in the 5th step of the path is now lower than it was with a 5 satellite horizon: the architects are modifying their decisions to attain more synergetic configurations later. Such higher posterior values do not compensate this loss from the optics of system 1, since they come heavily discounted.

In order to evaluate long-term scenarios and maximum attainable benefits, we can now show the evolution of synergy value across the path, as seen in Figure 51.

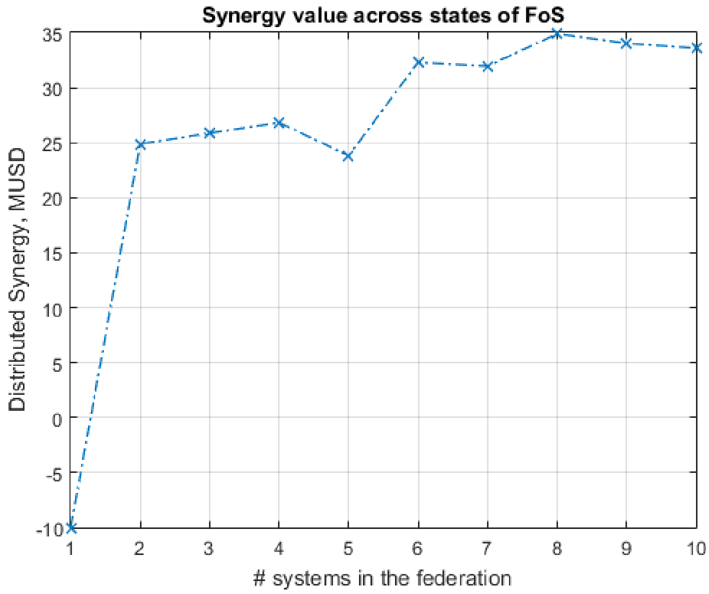

Figure 51. Synergy evolution through the commissioning of 10 satellite systems. 80 MUSD cost limit, 10 MUSD FSS cost assumption. The value of synergy is distributed amongst all systems, that is, we depict the average cost advantage for each through the successive FoS deployment states.

As explained above, the dip in the 5th system is cause by network topology limitations. Besides that, we can conclude the addition of the 2nd system is obviously the turning point in this case. From the 2nd system on, synergy grows and fluctuates in the range of 25-35 MUSD. Since the period of commissioning for the systems depicted in Figure 51 spans 10 years, and equals to the lifecycle of the systems involved, we can expect similar results when extending the analysis to larger amount of systems.

The fluctuation of synergy and the particular location of the 'dips' is connected, as we discussed, to case-specific network topology aspects. Nonetheless, this is general mode for the evolution of synergy due the change of decision-makers; driven by the openness or closeness of the federation and its governance mechanisms. For the 5th system above, to join the federation is locally optimal, and when considered in an isolated manner, it is also beneficial for the other systems to cooperate in such FoS configuration. However, previously commissioned systems might resent the slight decline in the distributed synergy present.

Several remarks need be made here. First, the distributed synergy is only an average estimator of the NBC for each system. Note that the incentives can vary how the synergy is distributed across systems' NBC, and hence can make the cooperation yields monotonically improving for the systems involved if the systems causing the dips yield most of their NBC in incentives to others.

Moreover, adopting new systems might allow the FoS to bridge to higher synergy scenarios as the results above show. Ultimately, the occurrence of such scenarios depends on the openness of the federation and the procedures to add new members. This phenomenon is also identified and studied in the WCN case study, in chapter 8.

6.5 Conclusions

This chapter has illustrated the different components of the framework and implementation details for a nominal case study, featuring 5 EO satellites in LEO orbit, and analyzed a series of additional cases. The demonstration of the framework in the 5-satellite set has shown that advantages of cooperating start with just 2 satellites, and that interface costs, architectural cost preferences, and network topology drive the achievable benefits. Specifically, when considering communication architectures under 80 MUSD for satellite systems, FSS interface lifecycle cost shall be kept below 15 MUSD to be an advantageous option. Under such assumptions, attainable RoIs are between 1 and 10 times the investment. When considering unrestricted cost assumptions, options with FSS interfaces are Pareto dominant even when FSS interface costs amount to 24 MUSD, and yield a RoI around 1.

This chapter exemplified the conditions of strong and weak synergy in a realistic case, the process of federating as an MDP, and the role of several key parameters of the framework. Several concluding remarks arise from the results here, both at framework and at case study level. We discuss now the role of FSS as compared to GR, the lessons learned from this case study and we close with methodological conclusions related to the framework.

6.5.1 The advantages of federating

For the cases highlighted in this chapter, federating leads to lifecycle cost benefits from a few MUSD to about 20 MUSD, for the missions involved. As mentioned, the value depends on the FSS costs, discount and sharing mechanism, network topology specifics and architectural cost constraints.

The results obtained in this chapter, under different assumptions, naturally lead to different cost advantages systems in the federation, and to different thresholds for federate/no-federate options. Nevertheless, from the results here we can extract characteristic values for the cost advantage of the first federate, as a ratio to the communications architecture total expense. For configuration path A, (Table 23) the cost advantages of federating equal to the 28% of the communications architecture budget. For configuration path B (Table 24), it is the 11.5%. For configuration path C (Table 25), gains are about 37% of the communications architecture expenses.

The value of federating in the cases depicted here, with bandwidth and latency chosen as performance metrics, origins from pooling the ground segments amongst federates and from the space network implemented, which effectively multiplies the opportunities for data downlink amongst federates. The underlying processes generating value on both cases are of different nature. In the case of bandwidth, note the conservation of data flows holds. We can change the allocations of bandwidth per federate via shuffling data between them, but the combined data downlink capability of a set of satellites is a constant once they are in orbit. Therefore, the value, the benefit of cooperation, on this case, is harnessed by re-allocating bandwidth from systems in a relatively flat bandwidth-utility response region to other systems that are in the verge of sharp utility increases in response of such extra bandwidth.

Bandwidth sharing generates weak synergy cooperation, analogous to what economy literature (section 4.1.4) called subadditive synergy: the ground station infrastructure cost is somewhat amortized by all the systems present, but no additional capabilities are created.

In the case of latency, the value instead arises by offering ad-hoc data relay opportunities, an option uniquely created by federation and the ISLs in place, and which has less tangible boundaries than the space-to-ground bandwidth. This is a superadditive, or what we called strong, type of synergy.

Table 32 summarizes the effects of the different parameters in the synergy and federation conditions and draws recommendations.

Table 32. Effects of different parameters on the prospects for federation.

Parameter	Effect on benefits	Example values run	Recommendations for federation emergence
FSS cost	↓↓	[1:1:24] MUSD	Best kept below 10-15 MUSD
Architectural cost limit Cc	↑	[20,30,80,∞] MUSD	Architectures constrained on their communications architecture expenses require of proportionally cheaper FSS interfaces
Network topology	↓	5 and 10 satellites in different orbits	Keep visibility to at least one federate
Discount rate	↓	[0.1:0.1:0.9]	A discount rate below 60-50% is more favorable
Sharing constraint	↓	[0.1,0.2,0.4,0.6,0.8,1]	Only slightly limits attainable benefits
Utility function	-	Nominal shape, alternative and heterogeneous (see section 6.4.1)	At least one system needs to be in need of additional performance (customer role)

The winning architectural paths feature a variety of datarates, but the options for ground segment strategy tend to be either 2 (two dedicated stations) or 4 (commercial lease). Only in the absence of cost constraints, the expensive option 5 is used. Options 1 (one station) and 3 (a more expensive commercial lease) are not in the SPF, but they were kept in the enumeration to study the appearance of non-Pareto baselines into the federated designs. As anticipated, neither options are found amongst federated paths. We can hence conclude that all of relevant federation options are built upon baseline architectures which are featured in the local SPF, adding the FSS interfaces on top.

The long-term analysis, extended to 10 satellites, shows a distributed synergy around 30-35 MUSD when the 10 systems are deployed, not very different from cases with less satellites. We can conclude the advantages of federation in this case are bounded by the network congestion and bandwidth ceilings that limit the exploitation of positive network externalities.

6.5.2 Geostationary Relay options as a substitute

As mentioned before, Geostationary Relay (GR) options do not appear in the federated paths as the combination of FSS and GR has been deliberately ruled out. The design impacts on pointing and acquiring simultaneous and non-compatible ISL systems, one of which requires significant power, and handling the data traffic as an effective communications hub make the case for something different than an EO spacecraft. Nevertheless, using a LEO spacecraft as a 'hub' to a GR service is an interesting architectural proposal that opens many more questions in platform design, service billing and the ability of the GR system to absorb such traffic. While not based on GR, orbital infrastructure services of this kind have been proposed (Palermo et al., 2015).

Nevertheless, GR options have been indeed used to build the SPF of each system, effectively comparing GR to other ground segment strategies and to federating. GR options, as implemented here, reach the highest latency utilities and moderate to high data rate capabilities (depending on the service, see Table 16). As such, architectures with GR are featured in the SPF, in the highest cost-utility region. However, as has been shown, such high performances can be achieved by federating at a significant cost savings. This stems from the high costs associated to the GR service. Taking as an example NASA's TDRSS, the most mature available GR service, we can perform some preliminary estimations here to support our case.

TDRSS features up to 6 geostationary operative satellites and additional spares (Frank J. Stocklin et al., 2012). While costs for different satellites and generations vary, we can safely consider the unit costs to be about 300 MUSD (Dan Leone, 2012). Let's consider for simplicity only the 6 operational satellites, procured simultaneously and with lifespan of 15 years. NASA declared the cost of TDRSS-M launch to be 132 MUSD in 2015 (NASA Kennedy, 2015). Let's assume accordingly a launch cost for unit of about 100 MUSD. This makes the procurement and launch for

6 units amounts to 2,400 MUSD. Yearly TDRSS operations cost are estimated at 85 MUSD (Martin, 2014). This is 1,275 MUSD in operations for 15 years, without considering depreciation.

In conclusion, before adding ground segment costs, the lifecycle costs of TDRSS are 3,675 MUSD. Per year, this is 245 MUSD. In 2014, 175,000 hours of mission support had been planned (Martin, 2014). This yields a cost of 23 USD per minute, which is the pricing that would amortize the investment in 15 years. This is close to the published minute rates of 21 USD for Multiple Access with return link (NASA, 2015b). For estimate of the Single Access link pricing, we need to consider a one-to-one ratio of TDRSS spacecraft to customers serviced. Hence, TRDSS can serve 6 missions with SA simultaneously. If the SA link is continuously used, in 15 years 47,304,000 minutes of service can be provided. Amortizing the cost on this service yields a pricing of 78 USD per minute, on the order of magnitude of the 132 USD rate announced. The latter is higher due the fact there are not enough customers to operate SA to maximum capacity constantly and thus, the infrastructure amortization is applied only to 60% of the theoretical minutes of service we assumed. Therefore, the minute rates of TRDSS and of any other GR services are tightly connected with the expenses incurred on deploying an expensive geostationary infrastructure.

LEO missions are in contact of their ground stations roughly 10% of their lifetime. Notwithstanding the potential disparity in data rates, we estimate the cost of using TDRSS for 10% during 10 years as 70 MUSD in the SA mode. The MA mode, with only 3 Mbit/s rate, is not a serious contender to the space-to-ground mission data downlink for missions with substantial on-board data generation.

In comparison to the 70 MUSD for the SA mode, the price tag for a fully-fledged, dedicated station capable of the same performances, is on the order of 25 MUSD (ESA, n.d). Leasing an existing set of stations costs less than a 1MUSD per year. Hence, comparing costs, GR is a poor choice for mission data downlink, as pointed in (Eilertsen, 2012). Even if we require additional latency, and more frequent contacts than 10% of lifetime, building or leasing additional ground stations nevertheless appears as a more cost-effective solution. Although real-life costs for FSS remain to be seen, the analysis performed in this section shows promise. While the usage of GR services involves the amortization of a dedicated infrastructure, in FSS the relay capabilities are piggybacked on the systems themselves, leading to significant cost advantages.

6.5.3 Theoretical insights

Besides the quantitative and qualitative insights for the applied case of FSS, multiple aspects of the implementation of the framework are of relevance for the method. We conclude here with remarks on the heuristic and tree search mechanisms, on the influence of the SPF topology in the baselines chosen for federation, the exemplification of weak and strong synergy, and the stability of the best architectural paths.

The test of greedy heuristics for MDP exploration and its comparison to the global optimal determines that synergy of intermediate states is a good estimator of final state achievable synergy. This a confirmation of an anticipated intuitive notion. Nonetheless, a one-step look-ahead of synergy in successor states is not enough until synergy starts being positive, i.e, from the second system in the case study here. Synergy is negative when systems are incurring in losses before experiencing benefits from federating; shall one encounter a FoS environment where this is the case for an extended period of time, the greedy policy shall be used with caution. The greedy policy is unable to adequately lead to the best future reward states if it needs to rank negative synergy states, especially when federates are not yet cooperating. For instance, consider a pair of systems, featuring a 10 MUSD federated interface each, albeit not using it, for instance due network topology issues. This state it has -20 MUSD synergy value regardless of the rest of architectural decisions, since those do not play a role yet in cooperation. Hence, the greedy policy is not sensitive to the rest of architectural decisions, which might kick in later when a third system federates to the previous. Yet, when starting from positive synergy and existing cooperation, it has been shown that the greedy policy can locate the global optimum or perform close to it.

Another interesting theoretical observation about the baseline architectures chosen can be learned from this exercise. In the absence of architectural cost constraints, that is, when the systems architects are free to choose amongst all Pareto-Optimal solutions, they favour high-cost and utility solutions in the federation environment. That is indeed the region were highest federating rewards can be attained. This follows from the morphology of the system's tradespace. Typical complex systems exhibit a diminishing returns area in their tradespace, that is, a stagnation of achievable utility when rising the cost to end of the range. This a commonly understood and accepted behavior of engineering systems utility-cost, or performance-cost relations, interpreted as the asymptotic approach to the highest attainable performances driven by physical limits (may that be material strength, optical aperture, solar array efficiency, etc). Hence in that high utility cost region, in the extra mile, an arbitrary increase of utility must have cost comparatively more investment that the same increase in the low-utility region.

By federating, systems' are experiencing such utility shifts, and by applying such shift in their high cost tradespace region, they are reach comparative cost advantages than if they adopt more modest design points. Despite this interesting fact, we were interested also to point out results in other interesting regions of the SPF, and thus enforced various cost limit constraints for the architects as detailed through the sections.

In addition, note the best architectural paths usually feature identical or very similar configurations through the parameter sweeps performed. That is, the best options for each system architect are robust to changes in discount rate, FSS interface cost, and also to moderate utility function changes.

Finally, the framework application has raised examples of both strong and weak synergy. While the latter is more common, strong synergy has been found in cases including low-cost architectures and low-cost FSS interfaces. Strong synergy in this case study is fundamentally connected with the improvement of latency, since it involves the idea of *'simultaneous additional performance for all'*. Bandwidth re-distribution alone can be weakly synergetic as described above, but not provide performance improvements for all parties since it's a constant, limited resource. Strong synergy instead requires of new value creation, or functional emergence, not simple resource re-allocation. The following case studies are examined in the light of these insights.

Chapter 7 Ridesourcing case study

7.1 Introduction

In Chapter 2, we briefly reviewed the principles of the sharing economy, defined as an access-based peer-to-peer market. Such markets present commonalities with FoS in terms of adoption challenge and peer dynamics, and provided inspiration for the advent and development of FSS. The differences between both, as analyzed in Chapter 2, are mostly to be found at the coordination level and the type of resources exchanged.

This chapter analyzes a specific access-based market, *ridesourcing*, and leverages on the methodology proposed for FoS to assess this example, outside of the engineering field. Through analysing the adoption of ridesourcing practices from a systems architecture perspective we demonstrate the general utility of the proposed approach. Moreover, since several ridesourcing services are in place, they can provide retrospective validation of the framework and its predictions. First, we need to briefly contextualize ridesourcing by defining the more general term ridesharing, and other commonly found terms, as carpooling.

7.1.1 The ridesharing context

Ridesharing is then conventionally defined as a *transportation arrangement where individual travellers split the expenses of a trip in a shared vehicle* (Furuhata et al., 2013). Hence, ridesharing is generally understood as a non-for profit activity for the parts involved; namely the driver and the rider. In order to provide for systematic opportunities to find riders and drivers, self-organized communities and also match-making agencies have emerged.

Within the umbrella concept of ridesharing, we find *carpooling* and *dynamic, or real-time ridesharing*. Carpooling is an activity that existed for decades: for instance, in the US it dates back to the 1940's (Chan and Shaheen, 2012). Back then, instead of using online postings or mobile applications, billboard listings were maintained within companies or business areas. Hence drivers and riders could plan in advance to share their daily commute, reducing costs, and, in more recent times, enabling them to use High Occupancy Vehicle (HOV) lanes. Modern carpooling works identically, but via online tools. The recurrence of the trip eases the planning efforts and the management of the cost and schedule expectations for the drivers and riders. For this reason, several carpooling services have already been trialled, studied or implemented in many cities

(Amey, 2010). Besides daily commuting, carpooling is also applicable to long-distance travel, which is generally planned in advance for both driver and rider parties. Carpooling can be monetized by match-making agencies via taking a rate of the transaction between drivers and riders, and such is the business model of BlaBlaCar (Shaheen et al., 2017), Covoiturage, Avego/Carma (Agatz et al., 2012), and similar companies.

Real-time, or dynamic ridesharing is a more complex scenario for match-making. In dynamic ridesharing, the trips are on-demand and do not feature long-term or advanced commitments between involved parties. Instead, the trips have an irregular schedule and Origin-Destination (OD) pair (Agatz et al., 2011). Dynamic ridesharing typically concerns urban, casual, short travel, as opposed to long-distance. Accordingly, it is subject to several modelling problems, especially in the case of real-time ridesharing with multiple passengers. Such problems include the minimization of combined travel costs and time, the appropriate matching of users, routing for multiple rider pickups and drop-offs, and reduction of associated detours. These problems have been studied from several perspectives and are still open (Agatz et al., 2012). Real-time ridesharing has, to date, not been successfully deployed at a large scale.

Figure 52, adapted from (Furuhata et al., 2013), represents the different modalities of ridesharing introduced here, using the trip OD nature (on-demand or fixed) and the drivers' motivation as classifying elements.

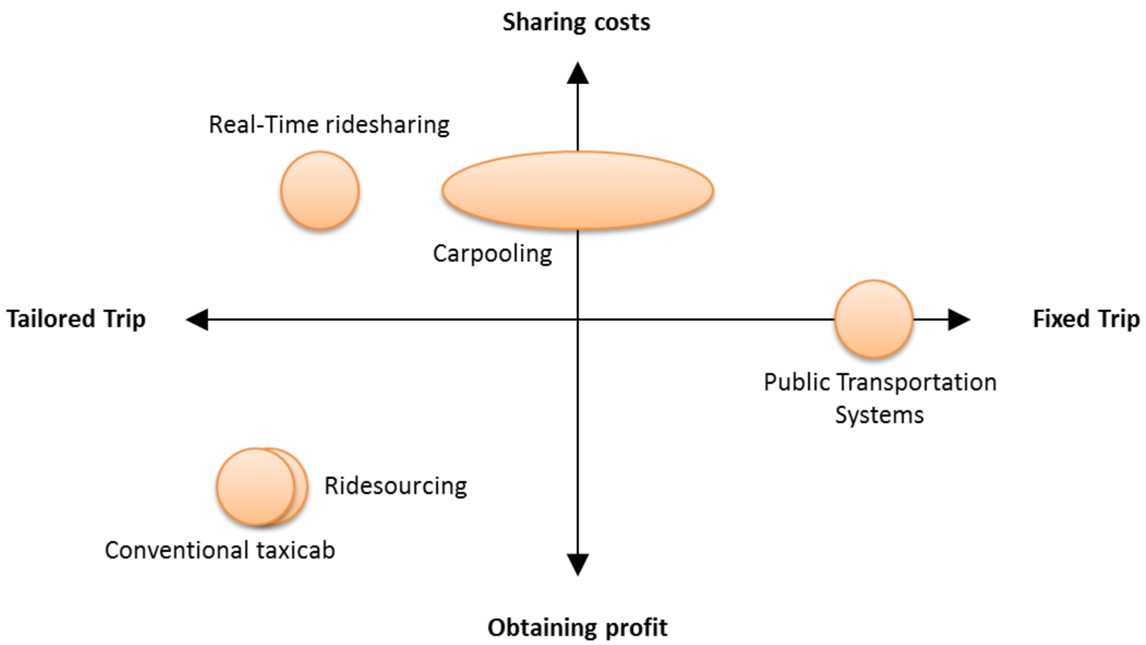

Figure 52. Ridesharing and ride-sourcing compared to conventional taxi industry and public transport systems, based upon the adjustment of the trip to the rider's OD demand, and the motivation of the driver or the service provider. Based upon (Furuhata et al., 2013).

In ridesharing, the driver's initial motivation to undertake the trip is not to offer a service, nor obtain benefit from the driving activity. The driver needs to go to a fixed destination due to her own personal transportation needs. In order to distinguish ridesharing from situations where the driver of the vehicle is motivated by profit and offers the trip as a professional endeavor, the terms *ridesourcing* (Zha et al., 2016) or *for-profit* ridesharing (Anderson, 2014) are instead used. The border, though, can be blurry; since the companies offering ridesourcing have been known to position their business as real-time ridesharing, and a fraction of the drivers working in ridesourcing are occasionally offering rides on the way to work or back home, as reported in the literature (Anderson, 2014).

The ridesourcing experience is in most aspects very similar to the classical taxicab industry, adding innovative dispatch interfaces that inform the user about expected fares, waiting time, and incoming car's position. As per today, several ride-sourcing services are available across the globe, including Uber, Lyft, or Didi (Harding et al., 2016).

A note apart is for carsharing, an activity that also has received renewed interest in the last years, due the advent of ubiquitous online connectivity and satellite-based location, which allows for de-centralization of the car fleet. In car-sharing, the users access a car without a driver, that is, they buy or rent a temporal access to an asset instead of a service. This has been the realm of traditional car rental companies, which have a fixed pickup and drop-off location for their cars. In contrast, the so-called Free-Float Car Sharing (FFCS) (Schulte and Voß, 2015) leverages on the mentioned technologies to enable dynamic pick-up and drop-off locations. In FFCS, users look for cars parked near their location using a mobile application, book them, drive them to their intended destination and drop them off.

7.1.1.1 Benefits and challenges of ridesharing

We can identify common challenges and benefits across different modalities of ridesharing. Benefits for the drivers are fundamentally of economic nature, both in for-profit and non-profit schemes. Driving new people can be also a socially stimulating and an incentive for long-trips, while this is a double edged sword and is also a part of the challenges described next. For users, ridesharing can be cheaper than its substitutes (such as taxi) or more convenient in schedule and flexibility, when compared to public transports.

Challenges of ridesharing (Amey et al., 2011) include adoption problems connected to the establishment of a critical mass, the safety concerns of sharing a drive with strangers, the mutual dependency of drivers and users, the difficulties of establishing a reliable service, the inconsistencies in vehicle type and driver/rider behavior, and the schedule tightness of some forms of carpooling.

In wider societal terms, ridesharing has been associated with traffic congestion and environmental advantages, as it puts in use the spare seats mobilized in private vehicles every day. However, this is only true in forms of ridesharing where the trip motivation responds to a driver need, and is not so clear in ridesourcing (Haider, 2015). Anderson differentiates both cases as *subtractive ridesharing*, where the amount of vehicles in the road is diminished by the activity, and *additive ridesharing*, where the net effect would be an increase of vehicles (Anderson, 2014).

7.1.2 Ridesourcing

On the remainder of this chapter, we focus on ridesourcing as the subject of the case study. As discussed in the previous section, ridesourcing is different from conventional ridesharing in *pricing scheme* and *driver motivation*. Ridesourcing companies control the ride pricing both at rider and driver ends, use a rate of the fare as a monetization strategy. As discussed above, the ridesourcing driver provides the vehicle and operates it usually as a professional endeavour, for profit.

The actors in ridesourcing are the driver, the rider or riders, and the coordination platform. The coordination platform matches drivers and riders, dispatching the former on demand. Coordination platforms are typically interfaced through a mobile application, where drivers and riders exchange location information. The rider is picked and dropped in on-demand locations, in the fashion of conventional taxi services. Figure 53 illustrates the general concept and information flows found in the ridesourcing market, notwithstanding differences across particular implementations.

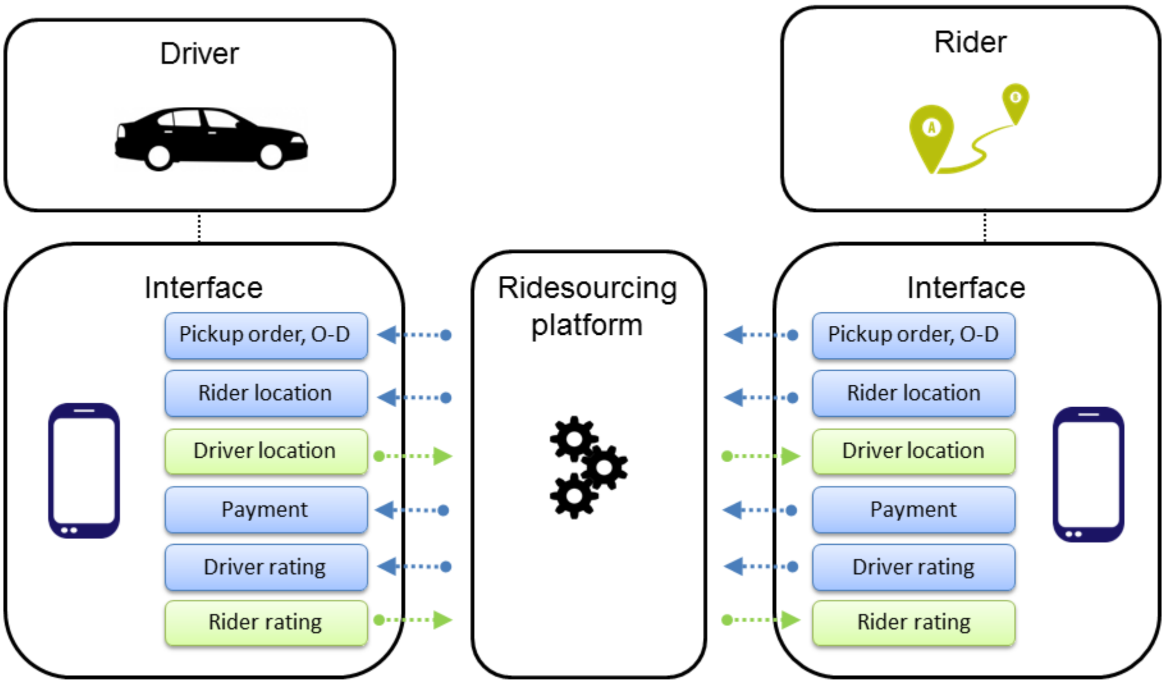

Figure 53. Typical information flows between driver and rider facilitated by the ridesoucing platforms.

Besides the OD pair, the location information and the payment processed through the platform, ridesourcing platforms routinely include driver and rider rating systems as many other online communities and businesses, as a ways to build trust, accountability and safety into the transactions.

7.1.2.1 The ridesourcing market

Since the appearance of the first ride-sourcing companies in 2009, they have enjoyed rapid expansion and success (Zha et al., 2016). The widespread adoption of smartphones across the population, and satellite-based positioning for the vehicles, remains the key technology enabler of ridesourcing and the reason for its introduction timeline. By 2015, Uber, the largest worldwide ride-sourcing platform by volume and capitalization, was generating more than 1 million rides per week, counting 3.8 million users, and operating in about 230 cities (Harding et al., 2016). At the time of writing this thesis, Uber is present in more than 500 cities worldwide (Uber, 2017), and has more than a 80% ride-sourcing market share in the US (Hartmans, 2016). About 20% of the US android devices have the Uber mobile app installed (Bloomberg News, 2016). Uber, while still smaller than the traditional taxicab industry, is catching up in number of daily rides across major US cities (Hall and Krueger, 2016; Schneider, 2016). In the US market, its main competitor is Lyft, operating in about 30 US states. Turning our eye now to Asian markets, it is estimated that 20% of android devices in India have the ride-sourcing application of Ola company installed (Bloomberg News, 2016). In China, Uber lost a ferocious market share battle to the local operator Didi, which is estimated to have 150 million users (Jon russell, 2015). Close competitors include Hailo and Gett, which offer a similar service, albeit do not fit under the category of ridesourcing. Hailo and Gett aggregate a fleet of conventional taxi services rather than enabling independent drivers to offer rides.

Nevertheless, the growth of ridesourcing has not been exempt of regulatory difficulties, sometimes leading to local bans (Haider, 2015). Ridesourcing operators have met widespread rejection and opposition from the taxicab industry (Fleisher, 2014), have inspired criticism in mainstream media for their aggressive business practices, and have been accused of shifting risks and labour expenses to drivers (Isaac, 2014). From a regulatory perspective, ridesourcing companies in the US have been classified under the novel legal figure of Transportation Network Companies (TNC), and not as transportation providers. This effectively spares them from the standard taxi industry regulations. The regulatory context is a key point of the ridesourcing business model, as several authors identify the ability to circumvent the licencing and taxi medallion system as the success driver for ridesourcing operators (Harding et al., 2016; Zha et al., 2016).

Moreover, the ability to summon and soak workforce as needed, without maintenance of driver employee contracts, is also a fundamental competitive advantage of ridesourcing companies against taxicab operators. The consequences for the taxi industry, the labour market, and public transport regulations have received significant attention in the literature (Chen and Sheldon, 2015; Isaac, 2014; Malhotra and Van Alstyne, 2014; Rogers, 2015; Slee, 2016), and are out of scope of the current work, which focuses on adoption aspects instead.

Ridesourcing companies are less affected by the ridesharing challenges mentioned in section 7.1.1.1. This is a result of professionalizing the driver pool, controlling the pricing, and heavily using driver subsidies (Newcomer, 2016). These measures stimulate a steady supply of drivers, incentivized to join the platform regardless of the amount of riders. Thus relieves the critical mass issues and the mutual driver-rider dependency. The professionalization of the drivers additionally allows for homogenizing the expectations on vehicles and behavior for the riders. As discussed before, the challenges related to safety concerns are mitigated through driver and rider feedback rating mechanisms.

7.1.2.2 Parallels with engineering FoS

Engineering and technical aspects of ridesourcing can be found at the interfaces and the coordination platform. In terms of technology, the specifics of the car assignment and dispatching algorithms, and the short-term dynamic pricing used in at least one platform are of academic interest (Cachon et al., 2016). Despite the obvious presence of technology as a key enabler, the components configuring the ridesourcing market are not engineering systems, and rather are socio-economical actors; the drivers and the riders.

Nevertheless, we can interpret the ridesourcing market in the light of the FoS definition, and argue that the ridesourcing market is a federation of drivers and riders. In Chapter 3, we defined FoS as *a set of engineering systems with independent goals, management and operations, that possess the adequate interfaces to cooperate and do so when it is advantageous for all parties involved.* Certainly the actors in the ridesourcing market have heterogeneous goals, management and operations if applicable, as they are autonomous individuals. In the case of riders, they choose their transportation mode based on perceived personal utility. In the same manner, drivers join the ridesourcing market based on individual benefit assessments and are their vehicles are not directly managed or operated by a central organization. Hence, for both riders and drivers, a clear value proposition must exist to join the market, and they can cease to operate in it, shall the benefits vanish. Even though the decisions faced by these stakeholders are not of engineering design nature, they indeed respond to cost-utility reasoning as in that case, and can be naturally modelled with the frameworks' techniques. The mutual benefit achieved by ridesourcing is what we termed synergy in engineering systems, and corresponds in this case to the notion of *social welfare* commonly found in economic literature, and also taxi market analyses (Zhang and Ukkusuri, 2016).

As in case of conventional FoS, a set of resources or capabilities are exchanged amongst peers, in this case, transportation for incentives. In ridesourcing, one of the peers is offering a service as her primary goal, which is not part of conventional federation concepts. Thus the role of incentives is critical and strong synergy is not expected; that is, it is not expected to be beneficial to the driver to offer a trip without incentives. The mobile interfaces and coordination platform common to ridesourcing are conceptually identical to the interfaces found in engineering FoS. However, the cost of adopting the interfaces in ridesourcing markets is deemed irrelevant compared to other cases like satellite federations, where constituent systems must piggyback their communication infrastructure.

Another different and thought-provoking aspect is the pricing control exerted by ridesourcing platforms, which allows for guidance and regulation of the amount of peers present in the platform. Hence, we find a certain degree of centralized oversight in this federation evolution. In the taxonomy of SoS, such centralization would, at first glance, incline us to think of directed SoS. Yet, ridesourcing platforms lack authoritative mechanisms to enforce the usage of the platform by the peers, hence respecting the Virtual SoS principles. Instead, the pricing controls are the way to influence the evolution of a federation, a feature that is not attainable by individual peers. Table 33 summarizes the interpretation of ridesourcing elements as FoS concepts.

Table 33. Application of FoS concepts to the ridesourcing market.

Conceptual element in FoS	**Interpretation in ridesourcing market**
Systems	Ridesourcing peers
Resource/capability exchanged	Point-to-point road vehicle transportation
Cost-utility of system alternatives	Cost-utility of engaging in the platform
Synergy	Social welfare
System’s NBC	Consumer and producer surplus
Architectural decisions	Pricing policies

7.2 Applying the framework

In the following, we apply the framework as described in chapter 3, illustrating all the specifics necessary of the ridesourcing case study. Ridesourcing poses the same challenges as engineering FoS in terms of adoption and network externalities. With the application of the framework to ridesourcing, we target specific goals. First and foremost, to study the deployment and adoption of ridesourcing including the multiple perspective of riders, drivers, and the platform pricing controls. So far, the literature has produced analyses of ridesourcing of static nature (Zha et al., 2016) and also preliminary assessments on the adoption dynamics of non-profit ridesharing (Agatz et al., 2011), but not an adoption analysis, including both the peers utility-cost reasoning and the pricing controls.

Additionally, we intend to capture the series of feasible equilibriums of rider demand and driver supply through the adoption of the platform, and predict the benefits for both rider and driver communities, the incentives necessary to encourage deployment in early phases, as well as the potential benefits for the ridesourcing platform. The case study will focus on the adoption of Uber services as a coordination platform, in the city of New York (NYC), from September 2012 to nowadays. The first step is to identify and characterize the FoS context. Table 34 captures all the notation that will be used used in this chapter. We maintain a cohesive nomenclature when possible with Zhang and Ukkusuri (2016), who studied the optimal fleet size for NYC taxi services under competition.

Table 34. Parameters used through this chapter, consistent with (Zhang and Ukkusuri, 2016).

Parameter	Symbol	Unit
Number of riders in ridesourcing	Nr	Riders/h
Number of drivers in ridesourcing	Nd	-
Average trip distance per passenger	Y	km
Price, rider side, ridesourcing	Pr	USD/km
Price, driver side, ridesourcing	Pd	USD/km
Demand	De	Trips/km/h
Demand intercept	a	Trips/km/h
Demand price elasticity	b	Trips/km/h/USD
Demand quantity elasticity	c	Trips/km/h/Car
Supply price elasticity	bb	Car/USD
Total road distance served	L	km
Average vehicle speed	V	km/h
Vehicle operation costs per unit distance	C_0	USD/km
Ridesourcing driver net earnings	B_d	USD/h
Taxicab driver net earnings	R	USD/h
Driver NBC (producer surplus)	ΔCd	USD
Rider NBC (consumer surplus)	ΔCr	USD
Synergy or social welfare	Sy	USD
Platform earnings	E	USD
Waiting time	Wt	min
Minimum wage	mw	USD/hr
Taxicab fares	Pt	USD/km

7.2.1 Identify FoS context

In 2011, Uber Inc. started a beta ridesourcing service in NYC, although until 2012 there is little data and evidence of traction available. By their own accounts, in September 2012 Uber had been operating for less than a year (Uber, 2014). We will take September 2012 as start of our analysis, and study the scenario until nowadays'. UberX service is the most commonly used of Uber products, a low-cost option based upon a 4-seater sedan and a driver accredited by Uber. From this point, our analysis refers more specifically to UberX.

UberX is a substitute good with respect to conventional taxis. Figure 54, adapted from (Schneider, 2016) shows the evidence of this. The pricing of UberX, below conventional taxicab, does indeed generate additional demand, but the major part of the UberX trips are substracted from the traditional the taxicab demand.

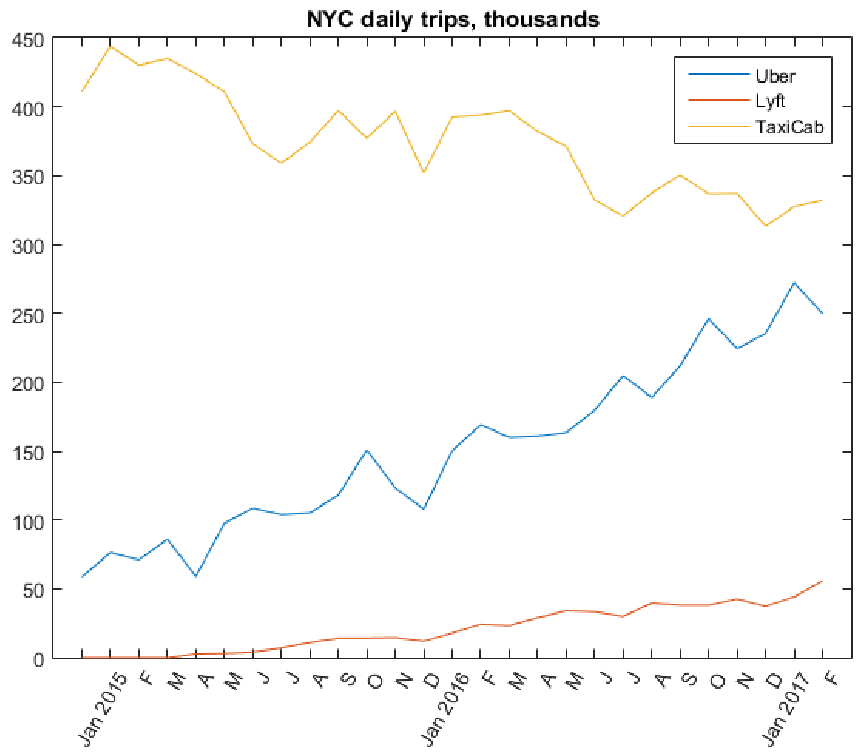

Figure 54. Statistics of daily trips provided in NYC by Uber and Lyft and conventional taxicab, adapted from (Schneider, 2016).

Figure 54 shows that the taxi demand in NYC is between 450,000 and 500,000 daily trips, a number consistent with the reports of the NYC Taxi and Limousine Comission (TLC) reports (New York City Taxi &Limousine Comission, 2014, 2016).

In NYC, as in many other cities worldwide, taxis are subject to a medallion system, or a permission per vehicle to operate, valued in the amount of thousands of USD. The price of this medallion peaked until close to 1 MUSD in 2013, and since then has declined back to 250 kUSD level, a circumstance analysts have precisely attributed to the advent of ridesourcing competitors

(Holodny, 2016). The amortization of the medallion adds to the taxi operation costs and eventually affects the service pricing, a factor effectively avoided by ridesourcing companies. Table 35 summarizes relevant figures of the NYC taxi market, as per 2013, just at the beginning of our analysis.

Table 35. Facts and Figures of NYC taxi market for 2013, extracted from the TLC yearly report (New York City Taxi &Limousine Comission, 2014) unless otherwise specified.

Parameter	Value
Vehicle fleet	13,437 cars
Trips	485,000 per day, (27,000 riders per hour in 2 shifts)
Typical shift time	9 – 9.5 h, 2 shifts per day
Driver wage	16.35 USD/h in 2016 (U.S. Bureau of Labor Statistics, 2016) 15.41 USD/h in 2013 (inflation adjusted)
Road network covered	12056 km (Zhang and Ukkusuri, 2016)
Average taxi speed	21.44 km/h (Zhang and Ukkusuri, 2016)
Average trip distance	4.18 km (Zhang and Ukkusuri, 2016)
Average fare	13.4 USD
Average fare per km	3.2 USD

An overview of the taxi market figures in other markets can be found in (Salanova et al., 2011). As can be seen in Table 35, a typical driver shift is about 9 or 9.5 hours. We follow here the approach of Zhang and Ukkusuri (2016), that concentrates all demand and operations in two shifts of 18h. Next, we model the decisions faced by the drivers and riders.

7.2.2 Formulate local utility and cost functions

Let us examine the utility-cost reasoning of drivers, riders, and the pricing controls that act as decisions. Unlike the FSS case, on which we had to use the aid of the SPF to map the utility advantages to cost units, in this case, the utility of federating can be readily connected to economic benefits. This illustrates a different technique for capturing federation advantages, as discussed in the chapter 5.

7.2.2.1 Rider cost-utility reasoning

The rider is a NYC city individual with OD pair and a range of possibilities at her disposal, including using all modes of public transportation, owning a private vehicle, hailing a taxi and of course, using ridesourcing. The topic of modelling the choice of transportation users has received wide attention in transport engineering literature. The seminal text of McFadden (1974) examines the selection amongst alternative transportation modes in the light of behavioral models and discusses the effects of a plethora of factors, including user income, personal preference, user demographics, and transport mode level of service (such as price, waiting time, travel time…). As introduced by McFadden, the basic model used in user choice modelling for transport is the logit model and its many extensions. The logit model is based upon conventional utility theory, and adds a stochastic error variable to the utility expectation of a particular choice by the user. This recognizes the inherent limitations of modelling human choices and the influencing factors. In virtue of the introduction of this stochastic error, when user(s) are confronted with two options we do not obtain a winning option, but a probability distribution of their choice. A systematic description of the practices of user choice modelling in transport engineering can be found in (Cascetta, 2009).

Nonetheless, the analysis of our case does not require of comparisons with all other transportation modes, since we assume ridesourcing is a substitute of the conventional taxi. In virtue of this, we can dismiss from our decision model the factors on which taxi and ridesourcing have an approximate parity (time of travel and comfort) and the demographics of the users (income, vehicle ownership, family status), and therefore dispense with the uncertainties associated to the modelling.

The premise here, supported by empirical evidence (Schneider, 2016), is that users of ridesourcing where a priori resolved to use the services of a taxicab, hence the other deciding factors are already pre-configured. What factors then affect the choice between a conventional taxicab and Uber? We select the price of the service, and the wait time as the key factors characterizing user experience in the taxi market (Wong et al., 2008; Yang and Wong, 1998). The price of the service captures the cost element for the federate, while the waiting time acts as the main utility measure. Thus we constrained the tradespace of the rider to two options, taxi or ridesource. Figure 55 illustrates this idea.

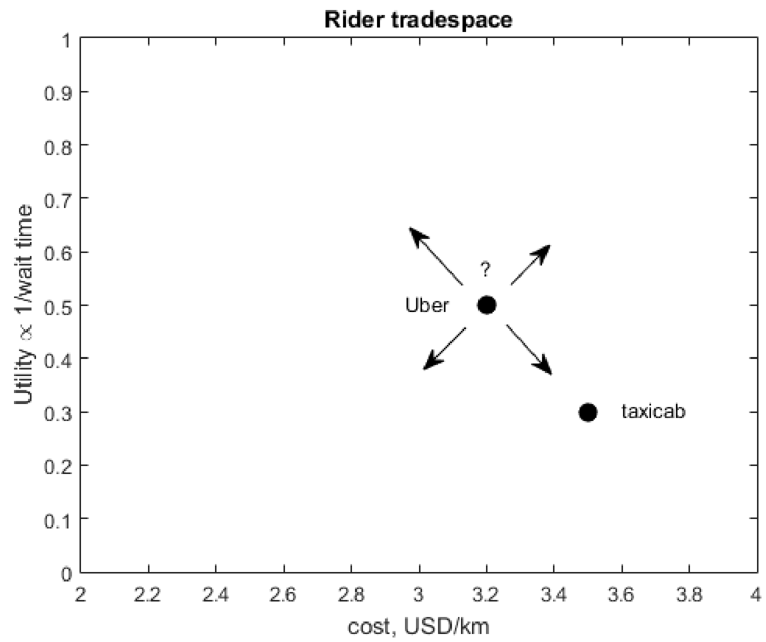

Figure 55. Rider tradespace reduced to two options, either use Uber or conventional Taxicab. Utility is inversely proportional to wait time.

As notionally illustrated, the cost and utility performances of choosing Uber depend on the other federates adoption and pricing decisions. Recall that in the FSS case study we were unable to account for utility in direct economic benefits and hence mapped it to cost through the SPF, a technique included in the framework. In the case at hand, we can map waiting time directly to economic benefits –or losses– using minimum wage as the conversion rate, a frequent technique in the literature (Zhang and Ukkusuri, 2016). Hence, the tradespace collapses to one dimension. We assume the rider will choose Uber if the NBC is positive as in Eq. 7-1:

$$\Delta Cr = (Pt - Pr)Y + mw(Wt_{taxicab} - Wt_{ridesource}) \qquad \text{Eq. 7-1}$$

Note Eq. 7-1 applies to the single trip, single rider. The performance of the standard NY taxicabs, in fares and waiting time, (Table 35) serves as reference point from which we establish the rider NBC. In classic economy terms, the NBC is roughly equivalent to the consumer surplus.

User base growth bounds

Before being confronted with the decision whether to adopt ridesourcing services, the rider needs to be aware of them. In order to capture this, we use the Bass diffusion model (Bass, 1969), widely used to forecast the rate of adoption of a new product. Researchers have already applied the Bass model to predict adoption of ridesharing services (Agatz et al., 2011).

The Bass model does not work well for products that are substitutes, and requires extensions to consider endogenous variables such as pricing (Bass et al., 1994). However, these factors are accounted for in the supply-demand formulations that will be introduced shortly, and we do not require the diffusion model to account for those. Instead, the diffusion model will act solely as an upper bound of ridesourcing users, limiting their amount to those who would be aware of the existence of ridesourcing services at a given time. The Bass diffusion process is commonly represented by Eq. 7-2:

$$\frac{f(t)}{1-F(t)} = p + qF(t) \qquad \text{Eq. 7-2}$$

Where t is time since product debut, $f(t)$ is the rate of addition of user fraction, $F(t)$ is the already installed user fraction, p is the so-called coefficient of innovation, and q is the coefficient of imitation. The innovation coefficient captures the external effects that make the product appealing, while the imitation coefficient q accounts for internal propagation effects such as word-of-mouth. In the absence of specific data, we adopt for them typical values of 0.01 for p and 0.3 for q. Eq. 7-2 can be re-arranged and applied in discrete timesteps to find the accumulated base fraction through time; we assume an absolute maximum user base for ridesourcing of roughly 30,000 riders per hour (Table 35). The time horizon for full awareness of the product through the market is assumed to be about 4 years. Figure 56 and Figure 57 show the evolution of product awareness up to 6 years after product launch.

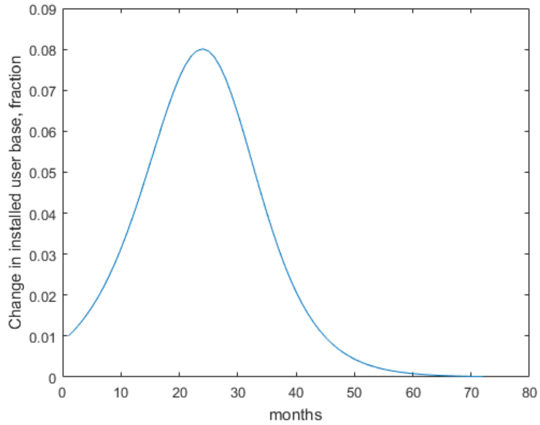

Figure 56. Change in installed user base fraction with p=0.01 and q=0.3.

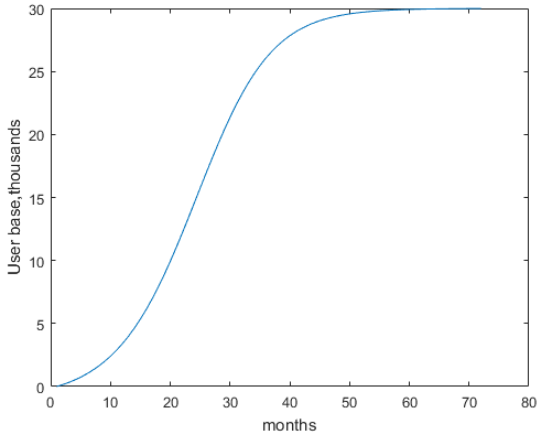

Figure 57. Evolution of the users aware of ridesourcing service, since product debut.

7.2.2.2 Driver cost-utility reasoning

The driver motivations to join ridesourcing can be safely reduced to considerations of economic nature. Surveys within the ridesourcing workforce support this point (Anderson, 2014; Hall and Krueger, 2016). We will assume two options for the driver: To work in a normal taxicab company with the associated wages, or to join ridesourcing. The evaluation of both options, as in the case of the driver, collapse to a single metric dimension. The two options can be compared to obtain the driver NBC:

$$\Delta Cd = Bd - R \qquad \text{Eq. 7-3}$$

The computation of terms in Eq. 7-3 and its connection to the number of riders will be introduced in the next section. Note that using *R*, the net earnings of conventional taxicab drivers, as decision threshold, is a conservative approach towards joining ridesharing. The lower barriers of entry might encourage drivers to drive for Uber even when the earnings are inferior to driving in a conventional taxicab company.

The adoption of drivers will also be bounded by a Bass diffusion process, with identical parameters as section 7.2.2.1 introduced.

7.2.3 Enumerate architectural decisions

In this case, the adoption decisions of the riders and users are not the only force behind the evolution of the federation; such adoption process can be influenced by an external actor, namely the coordination platform. The influence mechanism is the pricing strategy applied. We will model them as two architectural decisions: the price per km paid by riders and the price per km paid to drivers. The difference between both is the platform earnings *E*. Shall the price on the driver side be above than what the riders are paying, the platform is subsidizing the drivers and henceforth incurring in operational losses.

The decisions are implemented as increases and decreases in respect to the last pricing strategy; this rules out drastic pricing changes as unrealistic and not desirable from product positioning and adoption perspective. Table 36 lists the values selected as pricing controls. Note on the driver side the addition of a -5 USD decrease option, allowing the coordination platform to quickly recover from large initial driver subsidies.

Table 36. The pricing controls as architectural decisions faced by the coordination platform.

	Decision variables	**Feasibility bounds**
Pricing for user	Increase [-0.25, 0, 0.25] USD/km	Price not below 0.5, not above 3 (normal taxi 3.2)
Pricing for driver	Increase [-5, -1, 0 1] USD/km	Price not below 1, not above 20 (normal taxi 3.2)
Total decisions		12

7.2.4 Evaluate FoS configurations: compute synergy

The pricing decisions in Table 36 influence riders demand and drivers supply, changing the waiting times and NBC of both communities. In this case, a FoS state is defined by an amount of riders, drivers, and a price derived from the pricing control. Next we detail the modelling efforts to evaluate synergy and NBC for participants in each state. Figure 58 provides an overview of FoS state evaluation and the different models involved.

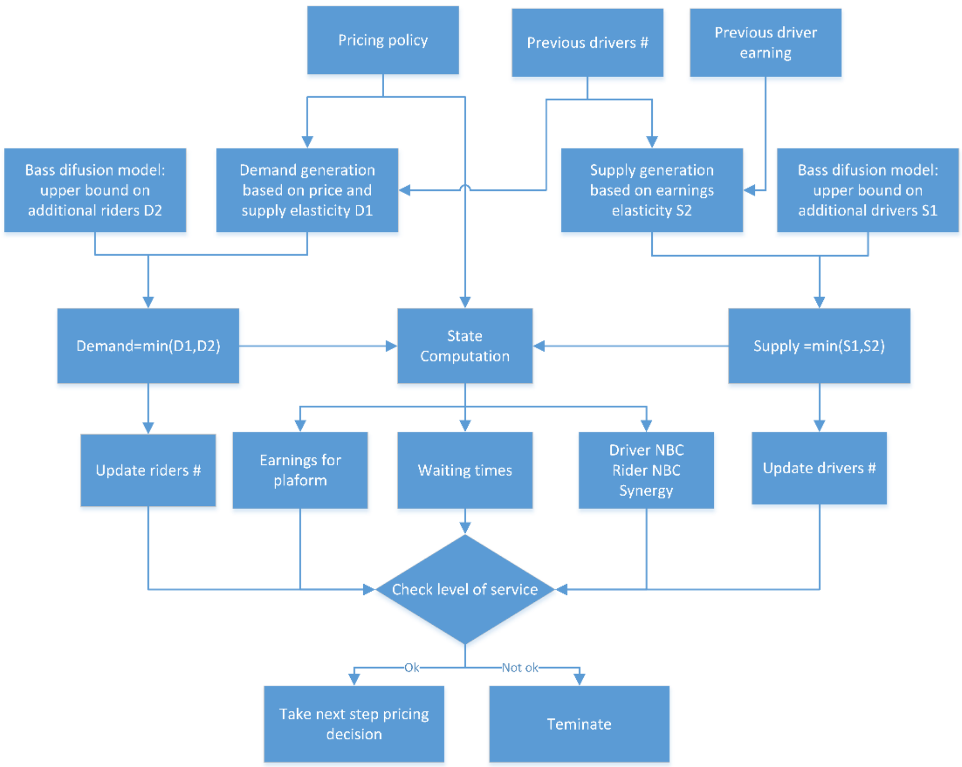

Figure 58. Overview of state evaluation procedures and involved variables.

While information of the previous state is necessary to evaluate the next, said information can be carried over to successor states with little overhead, respecting the Markov property. Besides the Bass diffusion models, already introduced, the state evaluation requires a wait time model, a demand generation model and a supply generation model.

7.2.4.1 Demand generation

The approach to the generation of the demand De follows the formulations of Zhang and Ukkusuri (2016). The demand is assumed elastic to the price and the amount of cars available, as expressed in Eq. 7-4.

$$De = a + bPr + cNd \qquad \text{Eq. 7-4}$$

Estimations of the elasticities b,c for the taxi market are available in the literature (Flores-Guri, 2003). Section 7.2.4.5 details the nature of these coefficients and the adjustment of the intercept a to the actual taxi market situation.

7.2.4.2 Supply generation

The supply is assumed to grow or shrink as a function of the earnings available. Data of the elasticity bb of the ridesourcing working pool to the available earnings is not readily available. However, we can proxy it with the elasticity of working hours of taxi drivers to earnings, for which studies in the taxi environment and also related to ridesharing exist (Chen and Sheldon, 2015; Sheldon, 2015). The growth in drivers, $Nd_{final} - Nd_{initial}$, is then the elasticity by the income difference with respect to conventional taxi wage. That is, for each 1% increase in earnings with respect to working in a conventional taxicab, the driver pool will grow in bb%. Eq. 7-5 expresses this concept.

$$Nd_{final} = Nd_{initial}\,(1 + bb\left(\frac{b_{d_{initial}} - R}{R}\right)) \qquad \text{Eq. 7-5}$$

Note that through the text we consider the amount of cars and drivers identical, that is, each driver has one car. Moreover, for the sake of the model we assume an archetype of ridesourcing driver that works in the same shift structure as typical taxis. Typical drivers in ridesourcing work significantly less hours than taxi drivers (Hall and Krueger, 2016). These discrepancies are accounted for and the values adjusted when comparing the framework predictions with real data.

Note on supply elasticity to earnings

The supply elasticity to earnings in the taxi market is a subject of controversy in the political economy literature. In an 1997 influential paper, Camerer and colleagues (1997) studied the behavior of NY taxicab drivers and surprisingly found negative elasticity to the earnings. This is not a common feature of the labor market; implying that taxicab drivers decide to finish their shift before the daily periods of more earnings. This justified a theory of 'income targeting' according to which taxicab drivers work with specific earing goals in mind and stop when such goals are reached, insensitive to the market demand. As reported by Sheldon (2015), this result has been proved and disproved a number of times. Furthermore, Farber attributed Camerer's results to econometric artefacts (Farber, 2005). Later, Chen and Sheldon studied specifically the case of Uber in ridesourcing, and found positive earnings elasticity (Chen and Sheldon, 2015). In this chapter we adopt this perspective.

7.2.4.3 Estimation of wait time

The wait time for the riders is a function of the city size and topology, and the available cars. We take here the formulation from Zamora, as reported in (Grau and Romeu, 2012) for rectangular-gridded cities (Eq. 7-6).

$$Tw = \frac{0.508}{V\sqrt{\frac{Nd}{L} - De \cdot \frac{Y}{V}}} \qquad \text{Eq. 7-6}$$

7.2.4.4 State Computation

We can now re-write the driver earnings term in Eq. 7-3, to obtain the NBC for a single driver, per hour, as:

$$\Delta Cd = \frac{De \cdot L \cdot Y \cdot Pd}{Nd} - Co \cdot V - R \qquad \text{Eq. 7-7}$$

For simplicity, the operation costs in Eq. 7-7 are computed as $Co \cdot V$, assuming the driver is cruising all the time, which is not necessarily true in the first stages of ridesourcing adoption, when the amount of users is small. This assumption is conservative towards the successful establishment of a federation. The NBC computation for riders stated in Eq. 7-1 remains the same, and we can combine the NBC from both sides to compute the Synergy in USD. Eq. 7-8 shows this computation. As usual, keeping Sy above 0 is a necessary condition for the establishment of the federation.

$$Sy = Nd \cdot \Delta Cd + Nr \cdot \Delta Cr \qquad \text{Eq. 7-8}$$

7.2.4.5 Model validation and calibration

The implementation of models described in this section has been validated against Zhang and Ukkusuri (2016), using the parameters of their study of optimal taxicab fleet for NY. The model developed here and theirs differ in the estimation of waiting time, and the exclusion of stochastic demand factors.

Some input parameters, such as the number of cars and cost is a result of the gaming approach of Zhang and Ukkusuri, but we use it as a validation point input for our modelling. Their work claims that the slight reduction of cars and fares would be socially beneficial for the NY taxicab market, which they shows to be oversupplied.

Table 37 Model validation against Zhang and Ukkusuri

Input parameter	**Value**	**Input parameter**	**Value**
Road network covered, *L*	12056 km	Car supply, *Nd*	10420
Car average speed, *V*	21.44	Car operation costs, *Co*	0.289 USD/km
Demand intercept, *a*	2.57 trip/h/km	Average trip distance, *Y*	4.4 km
Quantity elasticity, *c*	8.4e-5 Trip/h/km/cars	Price for customer, *Pr*	2.45 USD/km
Price elasticity, *b*	-0.5794 Trip/h/km/USD		
Output comparison			
	Model	Zhang and Ukkusuri	Reality (U.S. Bureau of Labor Statistics, 2016)
Earnings per hour, *Bd*	18.67 USD/h	19.01 USD/h	16.35 USD/h

The validation comparison is shown in Table 37. The results in driver income are acceptably close and validate the implementation herewith. Nonetheless, the data used by Zhang and Ukkusuri motivates a discussion about taxicab operation costs. The taxi operation costs of 0.289 USD/km, are based on data from Santiago de Chile (Zegras and Litman, 1997) and not NY, hence optimistic considering the mismatch in purchasing power. This is the reason for the larger earnings predicted by their implementation. For comparison, the American Automobile Association (AAA) calculated for 2014 an average of 0.366 USD/km for medium sedans (Stepp, 2014). For conventional taxicabs, due the medallion amortization, the number might be closer to 0.5 $/km. We now apply the model to reverse engineer the conventional taxicab operation costs and the waiting times that can be expected by them, as we need consistent estimates of the latter to asses Eq. 7-7. Table 38 shows the results for taxicab operation costs and wait time.

Table 38. Model reverse application with 2013 data to derive operation costs and wait time performances of NYC taxicabs.

Input parameters		
As in Table 35.		
Outputs		
Taxi operation costs	0.575 USD/km	Value that meets the average driver earnings
Wait time	2.21 min	Derived from the model

The value obtained for conventional taxicab operation costs is 0.575 USD/km. This value is not required to apply our model, since it is implicit in the net driver hourly average wage *R*. However the car operation costs for ridesourcing are explicitly required. Recent studies point to 0.37 USD per mile, or 0.23 USD/km for UberX drivers in NYC (Meyers, 2015), and that is the number adopted here for ridesourcing cars. These costs include vehicle ownership, depreciation, fuel, maintenance, tolls, repairs and insurance. This difference between taxicab and UberX operating costs, presumably originated from medallion amortization, directly affects the wages and prices and is the fundamental cause for the competitive advantage of UberX.

The demand elasticity has been estimated to be 0.72 to price and 0.47 to the available cars (Flores-Guri, 2003). With this information, the coefficients *a*,*b* and *c* can be adjusted to the realities of the 2013 taxi market. The supply elasticity is assumed to be 0.14 (Schneider, 2016). The minimum wage on the state of NY in 2013 was 7.25 USD/h. Table 39 summarizes all the parameter values used.

Table 39. Parameter values adopted.

Input parameter	**Value**	**Input parameter**	**Value**
Road network covered, L	12056 km	Bass model p	0.01
Average speed taxi, V	21.44 km/h	Bass model q	0.3
Demand intercept, a	2.79 trip/h/km	operation costs, Co	0.23 USD/km
Quantity elasticity, c	$8.08 \cdot 10^{-5}$ Trip/h/km/cars	Average trip distance, Y	4.4 km
Price elasticity, b	-0.503 Trip/h/km/USD	Taxicab earnings R	15.41 USD/h
Demand elasticity bb	0.14	Maximum wait time	30 min
Run start	September 2012	Taxicab fare Pt	3.2 USD/km
Run end	September 2017	Minimum wage Mw	7.25 USD/h
Pricing actions cadence	monthly	Taxicab wait time	2.21 min

7.2.4.6 Summary of modelling limitations and assumptions

The model to evaluate the states presented in this section is based on several assumptions and limitations that have been introduced through the text. We summarize them here for convenience.

The model assumes one passenger per ride, and the demand to be evenly distributed in an 18h service period. We assume UberX is a substitute of the taxicab services, with cross-elasticity approximately 1. We do not model the competition of other ridesourcing entrants as their market share has been proven to be small (Schneider, 2016).

The model for waiting times and demand-supply equilibrium is macroscopic, we do not include here considerations of microscopic nature such traffic network flow simulations and stochastic generation of OD pairs. This level of fidelity is enough to derive the output parameters of the model, and used in similar works (Grau and Romeu, 2012; Yang et al., 2000).

We do not include in the pricing schemes the tactical, short-period adjustments that the Uber platforms does, known as surge pricing (Cachon et al., 2016). Note that other ridesourcing companies, such as Lyft, do not engage in such pricing tactics, being those not an indispensable element of the resourcing market. We hence assume that dynamic pricing only induces short-term variance in the market parameters and the monthly effects on demand and supply can be determined by considering the monthly pricing average.

The model does not take into account external global factors such as traffic congestion and pollution derived from the variations in car activity, as we focus instead in the modelling of choices of individuals. Moreover, as the ridesourcing adoption is detrimental of the taxicab activity, we expect the numbers of vehicle in traffic to be approximately constant. Reports of the NYC mayor office (Office of the Mayor, City of New York, 2016) confirm that ridesourcing has not been, so far, cause of additional road congestion.

Variations in taxicab fares and number of cars through the 5 years of assessment have not been accounted for. Indeed, no large changes have been reported in fares and number of medallions (New York City Taxi &Limousine Comission, 2016, 2014). We neither contemplate the competitive reactions of the taxicab market to the ridesourcing incumbents. Such reactions are not apparent for fares, wages, or any other econometric parameters. However, actions at political level and regulations lobbying exist. Indeed, the TLC deployed in 2012 a new e-hail system for conventional taxis in a deliberate attempt to compete with ridesourcing companies. While difficult to account for, these issues might have significantly delayed the adoption of ridesourcing in NYC.

In this model, the behavior of the rider and drivers is explained purely by surplus and waiting time factors. This is naturally a simplification of human behavior and assumes perfect information. This parameter reduction is acceptable if we consider UberX a substitute to taxis and not a direct competitor to any other transportation mode.

7.2.5 Apply policy to select FoS configurations

We generate the states tree of this problem by applying the 12 possible pricing controls to an initial state, and recurring this operation, with a monthly cadence. States are evaluated on wait time, driver and rider NBC, synergy, and platform earnings as depicted in Figure 58. A monthly generation of successor states, through 5 years, leads to a tree depth of 60-levels. Taking into account the 12 architectural options, this yields a state space of potentially $6.1 \cdot 10^{64}$ states.

This tree is navigated from the perspective of the coordination platform and its profitability goals, hence the main driver of the MDP evaluation shall be maximizing the accumulated profit, while keeping the synergy above 0 and thus keeping riders and drivers engagement. A significant portion of this states represent undesirable situations, which are deemed 'game-over' and hence to be avoided by the MDP policy. These scenarios are *car shortage, large waiting times, negative user NBC, and drivers incurring in losses.*

7.2.5.1 Game-over scenarios

In Car shortage scenarios not all demand can be met, leaving users stranded. As this seriously damages user build-up, it is considered a failure for the coordination platform.

In a similar fashion, *large waiting times* –above 30 minutes as per Table 39– are assumed to lead to user disengagement and complete loss of revenue. While this is a recoverable situation, a well-planned ridesourcing operation should avoid this situation.

Negative user NBC refers to situations where, due the combined effect of waiting times and prices, using conventional taxicab is superior for the rider than ridesourcing. In such scenario, all of users switch to taxicab services and drivers earn no revenue. Again, while this is recoverable, the appearance of these states can significantly hinder user adoption and are to be avoided.

In states where *drivers are incurring in losses*, we assume they will completely stop operating, hence entirely collapsing the ridesourcing market. This is a less stringent condition than the positive NBC we enforced for users. Hence, we temporarily accept as valid the states where the drivers are having earnings, but negative NBC – they make money, but would be better working for a taxicab company-. Negative NBC will steadily decrease driver ranks as per Eq. 7-5, but they will not stop abruptly operating. This represents the driver lock-in, a factor less relevant for riders as switching service costs are little.

We avoid these states by assigning them infinite losses for the platform. In this manner, we encourage the MDP policy to only select sustainable states. The MDP policy to explore the tree is discussed next.

7.2.5.2 Policy proposal

In order to navigate the state space, we will apply a MDP policy based upon maximization of platform income. The MDP policy adopted is to pick the successor state with maximum reward *Re*, of the form:

$$Re = \frac{k_1 \cdot |E|}{k_1+k_2+k_3} + \frac{k_2 \cdot |\Delta Cd|}{k_1+k_2+k_3} + \frac{k_3 \cdot |\Delta Cr|}{k_1+k_2+k_3} \qquad \text{Eq. 7-9}$$

The Earnings *E*, and NBCs in Eq. 7-9 can be normalized based upon the maximum numbers of riders and users, using the NYC taxicab market an upper bound. These formulations aims to favour the choosing of higher earnings states, while keeping riders and drivers engaged in the platform. The values k_1, k_2 and k_3 have been set through trial and error. Forty different policies with different combinations of the coefficients have been tested. Most of such trials ended either in game-over situations or in platform losses, both undesirable end states. Some observations from these tests are summarized here.

The factor k_3 can be set to 0 as the engagement of the riders is a result of the competing forces of driver engagement and platform earnings. In plain words, if drivers are kept happy, and yet the platform is accumulating earnings, this renders acceptable pricing and waiting times for users. However, setting also k_2 to 0 and prioritizing only the maximization of platform earnings naturally leads to disaster in a short-sighted search, as greedy pricing constricts any necessary room for growth. The best settings found for the coefficients are 0.8, 0.2 and 0 for k_1, k_2 and k_3, respectively. Different coefficient values in the range of ±0.05 of the proposed ones also yielded platform earnings, of a lesser amount. This policy does not guarantee to find the optimal earnings accumulated after the 5 years of MDP, but does yield a feasible solution arguably close to the real optimal.

7.3 Results

We now proceed to apply the proposed policy to explore this cases' MDP states tree, evaluating each state with the techniques. The initial state, in September 2012, is adjusted to be marginally beneficial for both riders and users, under the conditions summarized in Table 40.

Table 40. Initial state conditions

Parameter	**Value**	**Parameter**	**Value**
Initial number of riders	14 riders/h	Initial number of cars	700 cars
Price for user	0.5 USD (min)	Price for driver	20 USD (max)

The initial conditions are derived from the minimum amount of cars necessary to meet the waiting constraint of 30 minutes, under minimal demand. This is 609 cars. In order for these cars to operate with benefits, under the maximum subsidies allowed in our formulation (19.5 USD/km), they need to attend a minimal demand of 250 riders per day, or 14 riders/h spread through an 18h service period. An alternative to this is to simply pay them fixed hourly wages irrespectively of the existence of a demand. Ridesourcing companies employ this strategy to kick-start their business. In order to avoid tailoring the formulation to this initialization specific, we instead opt for assuming a modest initial demand.

After assessing the MDP with the proposed policy, we obtain a configuration path of 60 states, one per month. On the following figures, we track the synergy or social welfare, platform earnings, accumulated earnings, waiting times, pricing actions undertaken, and driver and rider adoption, through the path's states.

7.3.1 Synergy and platform earnings

Figure 59 shows the synergy, or social welfare, and the platform earnings, month by month. Due the open-loop action of the pricing controls and the effect of rider and user adoption, both quantities oscillate until the 40th month where they reach an equilibrium. Until the 18th month, the coordination platform is incurring in losses due driver subsidization. From then, the amount of riders is enough to provide for drivers' incomes. Social welfare and platform earnings are, as expected, negatively correlated.

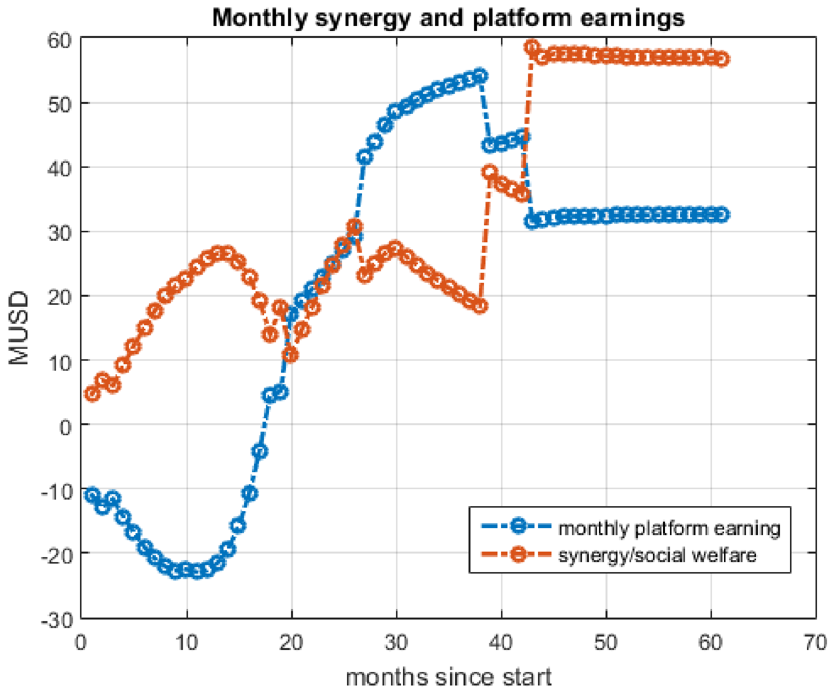

Figure 59. 5-year evolution of synergy and coordination platform earnings month by month.

Notably, on the last stage the platform is providing the largest level of social welfare to riders and users, but not the largest earnings for the platform. The second issue is due an artificial peak of earnings which takes place between months 27th and 38th. This is attributed to the observed decision to increase the user pricing as the market reaches maturity in the 27th month. This stalls user growth and leads to a re-consideration of user pricing after the 38th month, as discussed in the next section.

Figure 60 shows the platform accumulated income, resulting from the depicted monthly earnings. Before starting to generate positive cash flow on the 18th month, upfront losses total 290 MUSD. After that, earnings steadily accumulated at slightly varying rates, subject to the pricing variations. After 5 years of operations, the platforms earnings are about 1230 MUSD, or about 246 MUSD per year. The time to break-even is about 30 months.

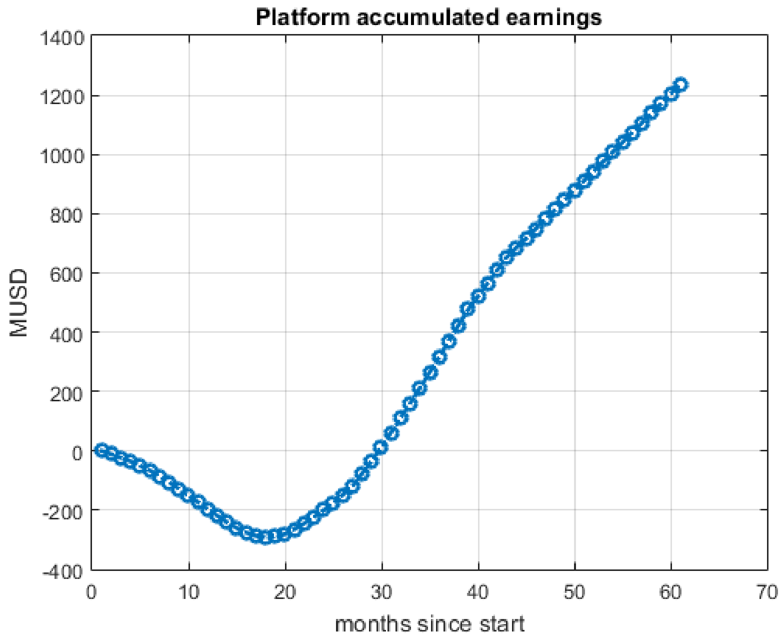

Figure 60. Accumulated earnings through 5 years of operations for the ridesourcing platform.

7.3.2 Pricing controls

The decisions to increase or reduce prices acts as an open loop control, with feedback delayed at least a month. Figure 61 illustrates the evolution of the prices, as different pricing actions are undertaken each month.

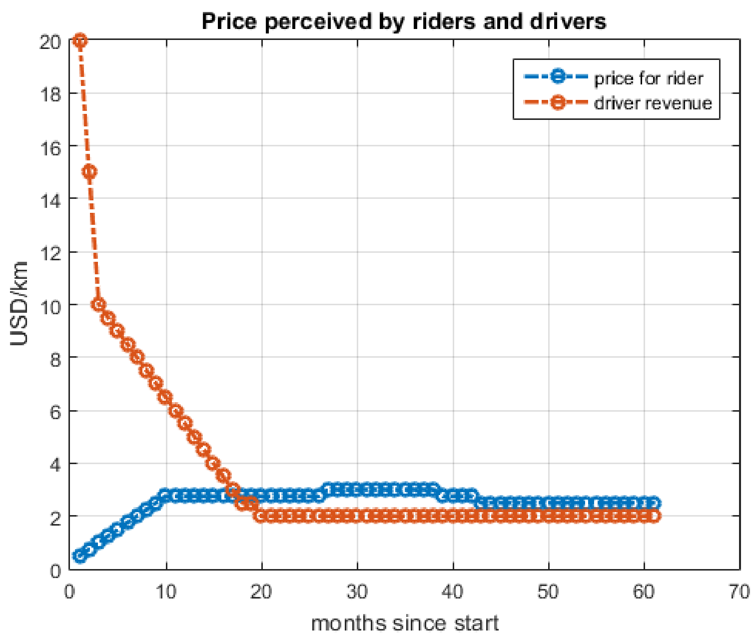

Figure 61. 5-year evolution of the price paid for a ride by the passenger, and compensation received by the driver.

Driver returns are above rider payments until the 18th month. The driver compensation per km drops steadily in the first 2 months, from the start at 20 USD/km to 10 USD/km, with the platform using the -5 USD/km control. From then, it drops at a steady rate until settling in 2 USD/km. On the side of rider pricing, the platform attempts several strategies including 3, 2.75 and 2.5 USD/km.

From the 27th month, the user base and waiting times –shown next– are mature enough that slight user pricing increases do not have drastic negative effects future in income returns, as demand and supply are high enough. Moreover, the growing mismatch between drivers and riders requires to reduce the adoption rate of the latter, or risk approaching game-over scenarios. Guided by these premises, the policy selects such user pricing increases. These behaviors stall the demand, and ultimately require from lowering again the prices as happens in the 38th month. Not much later, the amount of riders and drivers hit the market ceilings set by the Bass model and the pricing scheme stabilizes. The final pricing is 2 USD/km for drivers and 2.5 USD/km for riders, which represents a commission of 20% over the rider fare rate, in line with Uber data (Haider, 2015).

7.3.3 Riders and drivers adoption

The amount of drivers and riders increases steadily through the MDP states. The effects of the pricing discussed in the previous section lead to the rider adoption stall and recover around the 38th month, as shown in Figure 62. Through the 5-years MDP exploration, the rider adoption is mostly bounded by Bass diffusion effects except on the aforementioned demand stall periods. The driver adoption is instead controlled by the supply elasticity.

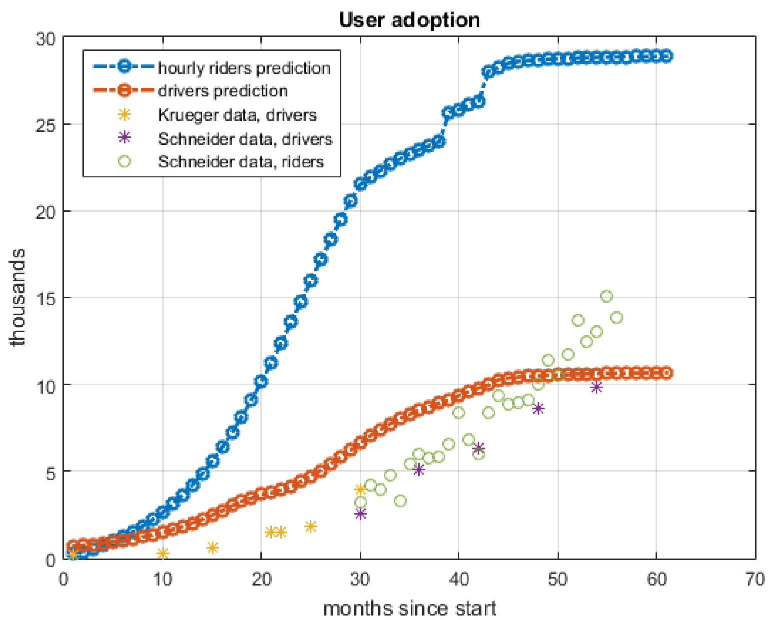

Figure 62. Predicted evolution of rider and driver adoption in 5 years, compared with data points from Schneider and Krueger. Note the driver data needs to be adjusted taken into account the working hours of ridesourcing drivers.

For reference, we include data from Krueger and Schneider (Hall and Krueger, 2016; Schneider, 2016). For the sake of comparison, the rider trips data from Schneider needs to be divided in an 18h service period, as our data is counted in riders/h. The driver data refers to the amount of simultaneous cars on the street, and hence to be corrected computing the average working hours of ridesourcing drivers. Based on the surveys found in (Hall and Krueger, 2016), one derives the average NY Uber driver works about 22.1h per week, roughly half of a normal taxicab driver. Taking into account 2 shifts of 18h, we conclude for every 1000 drivers registered in the Uber platform, there are about 245 cars active on the street during service periods.

The predicted numbers for drivers are consistent in evolution with the available data and reach similar values at the end of the analysis period, with about 10,000 drivers active at any given time. However, the numbers of riders have a significant mismatch, being the prediction 2-3 times larger than what Schneider data shows. This delay in user adoption originates from taxicab market reaction and behavioral aspects of the user choice that have not been modelled. Section 7.3.5 delves into this topic and other validation comparisons.

7.3.4 Waiting time

Figure 63 illustrates the evolution of waiting time for riders. The waiting time suffers a monotonic reduction, only interrupted between months 18th and 27th due rapid increase in users. The waiting time settles in a value of 3.4 minutes after 40 months.

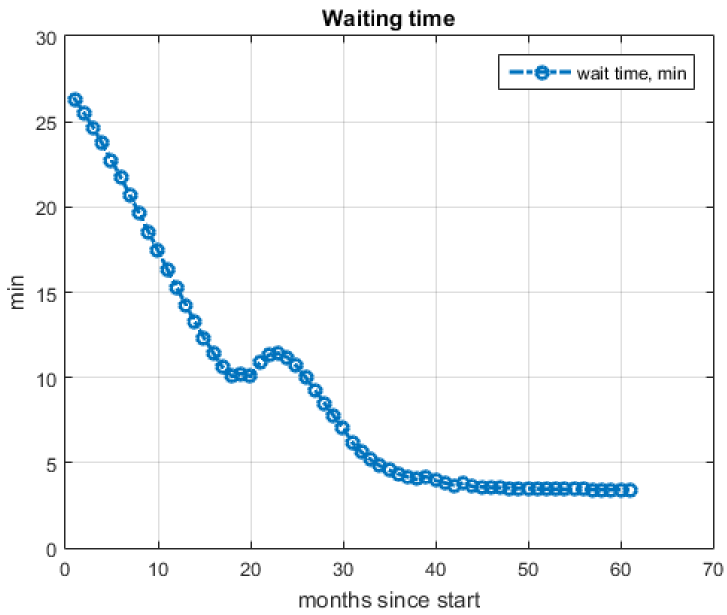

Figure 63. Evolution of predicted waiting time.

7.3.5 Validation discussion

The comparison of the results above and the data publicly available, yields, with all due caution, a good agreement between the framework predictions and unfolding of events in the NYC ridesourcing market. Table 41 compares several key values with data from other sources, when available.

Table 41. Numerical validation results, for 5-year framework predictions.

	NYC Data	Framework Prediction	Error
Price for rider	2.68 USD/km (Uber Estimate, 2017)	2.5 USD/km	-6.7%
Driver earnings	Approx. 21 USD/h (Hall and Krueger, 2016)	17.6 USD/h	-16.1%
Platform commission	Approx. 20-25%	20%	-5%
Time to breakeven	Not publicly avail.	30 months	-
Time to first earnings	Not publicly avail.	18 months	-
Upfront expansion losses	Not publicly avail.	290 MUSD	-
2016 wait time	2.25 min (Alley, 2016)	3.4 min	51%
2016 earnings	Approx. 400 MUSD (Alley, 2016)	388 MUSD	3%
March 2017 number of riders	15,276 riders/h (Schneider)	28,817 riders/h	86%
March 2017 number of drivers	10,608 drivers/h(Schneider)	10,626 drivers/h	0.2%

The only parameter featuring a large difference between estimates and market reality is the amount of riders. Our model predicts that by 2016 UberX would have substituted almost entirely the conventional taxicab services. This process is clearly underway, but only approximately half-way (Schneider, 2016). Undoubtedly, this difference stems from the lobbying and modernization actions undertaken by the taxicab industry (Office of the Mayor, City of New York, 2016), which is a significant source of income to the city municipality. Moreover, our rider decision model is unable to capture the effects of public attitude towards Uber as a brand (Dawes, 2016). To a lesser degree, this mismatch can be also attributed to the competitive pressure of other companies like Lyft, Juno, or Gett.

The waiting time estimates in the literature and media differ widely (Alley, 2016; Sender, 2014; Uber, 2014), and also are different across NYC boroughs. The precision of this value is naturally limited by the specifics of our macroscopic approach. However, a difference of 1.15 minutes is deemed acceptable. The predictions and data available for fares and earnings are remarkably close, as is the prediction for the fare commissions. Being Uber not a publicly traded company, detailed data on their NYC expansion losses and time to breakeven is not available. Nonetheless, estimations are available for the 2016 platform earnings, for which we predict a fairly close value. Finally, the number of NYC drivers engaged Uber as per March 2017 is also remarkably close to our estimates.

7.4 Conclusions

This case study on the NYC ridesourcing market, from September 2012 to September 2017, has illustrated the application of the framework to federations outside of the engineering scope. The UberX ridesourcing service has been operative for more than 5 years in NYC, providing a validation of the framework estimates. This study provides specific insights to the ridesourcing market, and also general insights related to the methodology.

7.4.1 Case study insights

The application of the framework identifies positive conditions for the federation emergence when adoption numbers total, as shown by the figures, around 10,000 riders/h and 4,000 drivers. At that stage, the federation is self-sustaining, both riders and drivers benefit, and even earnings for the coordination platform can be obtained. The benefits reported here, totaling about 55 MUSD/month (Figure 59) are consumer and producer surpluses with respects to the conventional taxicab service.

Before the self-sustaining stage, it is necessary to subsidize driver operations. This, by our accounts, leads to losses in the amount of 290 MUSD before the start of positive cash-flow. While the exact data by cities is not available, Uber has been frequently reporting losses worldwide on the amount of 2 Billion USD per year (Newcomer, 2016). The company attributes them, as we expected, to the driver subsidies needed for global expansion. For reference, with the data presented here, these losses would fund the expansion into 7 cities like NYC per year, or the equivalent in smaller cities. Uber has a very important venture capital backing, and it needs it for its quick-paced and aggressive expansion strategy.

Provided the user base has been built, why is it beneficial both for riders and drivers to adopt ridesourcing as compared to taxicabs? The single most important reason is the cost of car operation. The costs of operation of UberX cars, estimated in 0.23 USD/km, are less than double the operation costs of NY taxicabs. Two opposite effects can be distinguished to argue about the operation costs. On one hand, the individual drivers of Uber do not experience any economies of scale. In atomizing the workforce, Uber cannot make use of the cars as intensely as taxicabs, neither benefit from car commonalities in repairs, and any other fleet advantages in parking, or insurance, for instance.

On the other hand, the amortization of the city medallion is an additional cost not experienced by Uber drivers. This effect appears much larger than the loss of economies of scale. The medallion value can widely change every year, but, let us, as a thought exercise, pick a medallion cost of 500,000 USD, consistent with market values across the 2000's. U.S tax regulations assume a period of amortization for medallions of 15 years, and in any case not more than 25 years. A typical taxi makes about 70,000 miles a year (New York City Taxi &Limousine

Comission, 2014). Hence, if we pick a 20 years amortization period, before inflation and depreciation adjustments, this yields a charge per kilometer of 0.22 USD. This justifies the observed differences in car operation costs (Table 38, Table 39).

As discussed in section 7.3.5, our framework, due modelling limitations, does overestimate the number of riders. However, the platform earnings are close to real data. The reason for this is Uber's dynamic pricing. While through this text we use averages, Uber demand is typically more concentrated in time than taxicab. The use of dynamic pricing allows for leveraging this, as increases the fares and earnings. Hence, while the expectation of the NYC UberX fare *through all day* is 2-3 USD/km, at the times the demand is concentrated the fares are larger, compensating for the effect of the relatively smaller demand (Cachon et al., 2016).

Dynamic pricing also attracts more drivers to the road when demand peaks and increases the car occupancy rate with respects to conventional taxicab. These effects yields Uber more effective in producing revenue than conventional taxicab services (Alley, 2016).

7.4.2 Theoretical insights

Besides the case-specific results, we can extract relevant methodological observations from the application of the framework to a ridesourcing federations, and compare it to the previous case study.

In the case of ridesourcing, it takes about 14,000 individual decisions to achieve a sustainable, beneficial federation, while for satellites, 2 or 3 systems were enough to establish positive synergy. In these two cases, the weight of the decisions of particular federates in the success of the federation is very different. Convincing 14,000 individuals to sustain economic losses for 1.5 years to establish a beneficial ridesourcing community, is, at best, challenging. For casual rider it is not common to think of taxi services in a 5-year horizon; and drivers wouldn't engage for hypothetic future benefits in a market without barriers of entry. That is, the long-term returns of establishing the federation are not apparent in the local decision spaces of the first potential ridesourcing federates.

Furthemore, the concurrent growth of riders and drivers in the federation needs to be closely managed to reduce losses through the adoption period, and to keep a sustainable equilibrium after. Borrowing the notion from Maier (1998), *intermediate stable forms* are desirable through the FoS build-up.

For all these reasons, the role of incentives in this case plays a bigger role; as only an external entity is able to supply the incentives and strategize the federation deployment with long-term perspective. We can find actors capable of this either in the governmental side or in the private sector, as Uber Inc. In any case, significant upfront investment is needed to achieve rapid expansion. In comparison, the role of these entities could be applied in a peer-to-peer fashion in the FSS case study, as indeed, satellite manufacturers and stakeholders are in a position to plan in long term and capture the returns of their federate investments.

Finally, note how in this case we exemplified a different mechanism for capturing the utility and cost reasoning of federates. The utility for drivers and riders can be readily captured as economic benefits, dispensing the utility-cost SPF mapping techniques employed in other cases. Despite the differences in the modelling, this chapter successfully generalized the framework premises and FoS concepts to federations of socio-economical nature.

Chapter 8 Rural Community Wireless Networks

8.1 Introduction

Wireless Community Networks (WCN) (Oliver et al., 2010) are meshed networks based upon wireless links between particulars, independent from established Internet Service Providers (ISP) and operators. Notable WCNs include *Guifi.net* (Vega et al., 2012), *Freifunk* (Herberg et al., 2010) and the *Athens Wireless network* (AWMN, 2016), amongst others.

WCNs are peer-to-peer private networks, not including access to the worldwide internet per default. However, participating into a WCN can grant internet access through peer proxies. Such proxies, called here Access Points (AP) can resell or share bandwidth. Hence, WCNs can also provide internet access to the users residing in areas with poor connectivity, therefore being of particular interest in rural or sparsely populated areas.

This chapter characterizes rural WCNs as federations and analyzes their adoption dynamics, taking into account demographics, network topology, and alternative internet provision services. In particular, we focus our study on rural areas of eastern Spain, as one of the world's major WCNs, Guifi.net, was conceived there.

8.1.1 The rural digital divide

Rural areas worldwide suffer from low penetration of internet access respect to urbanized regions, especially in the broadband segment. Broadband internet is loosely understood as a download speed in the order of single-digit Mbit/s. The US Federal Communications Commission (FCC) defines the threshold of broadband at 4 Mbit/s for streaming and teleconferencing services, while other sources use 1 or 2 Mbit/s threshold (Federal Communications Provision, 2017).

Such levels of performance still have not reached rural users worldwide. In the UK, from 12% to 47% of rural households have no broadband access or no internet provision at all, depending on the sparseness of households and remoteness of the area (Townsend et al., 2013). In the US, 16% of rural households have no broadband access (Kuttner, 2012). In the EU, 9% of rural households don't have internet provision, 17% have no access to low-speed Digital Subscriber Lines

(DSL) based on traditional telephone copper wire, and 72.2% experience speeds lower than 30 Mbit/s (European Comission, 2016). Table 42 introduces a brief overview of the technologies discussed here. The 30 Mbit/s speed level is a priority target of the EU's Digital Agenda (DA) which aims to have *more than 30 Mbit/s coverage for all by 2020*, using a mix of technologies, both wired and wireless (European Comission, 2013). In countries with less overall internet penetration, the access problems for rural population are naturally more acute. For reference, note African countries exhibit only a 25 % of overall internet penetration, the Arab states exhibit a 42%, and India a 35% (ITU, 2016).

Table 42. Internet access technology definitions, adapted from (European Comission, 2016).

	Technology	Details	Typical performances
Fixed	DSL (Digital Subscriber Line)	DSL, and asymetric DSL (ADSL) protocols serve broadband over conventional telephone lines	Typical about 4 Mbit/s, drops below 1 Mbit/s over 5 km line
Fixed	VDSL (Very-high-bit-rate Digital Subscriber Line)	Combines street cabinets equipped with fibre backhaul with telephone lines for the last mile	Widely varying, depending on distances involved
Fixed	FTTx (Fiber-to-the-x)	As opposed to VDSL, FTTx deploys fiber optics closer to the users, either to the curb (FTTC), the client premises (FFTP) or home (FTTH).	100 Mbit/s to Gbit/s
Fixed	Cable Modem	Uses coaxial cable, such the one for TV	Typical 20 Mbit/s
Fixed	DOCSIS 3.0 (Data Over Cable Service Interface Specification)	The DOCSIS 3.0 standard uses also coaxial cable, but achieves larger speeds than conventional cable modem	Typical 30 Mbit/s
Wireless	LTE (Long Term Evolution)	Mobile broadband service based upon licenced spectrum frequency allocations, in the 700-2600 Mhz range	30-100 Mbit/s, depending on distance
Wireless	WiMAX (Worldwide Interoperability for Microwave Access)	Wireless standard using unlicensed bands in the 2.4 and 5 Ghz spectrum portions	30 to 500 Mbit/s

This difference in internet penetration and broadband availability between urbanized and rural areas, oftentimes termed *the digital divide* (Rogers, 2001), hampers economic development and welfare of rural regions (Kuttner, 2012). The access to broadband has positive impacts across many activities, including education, healthcare, job creation, e-commerce, and many others (Evangelista et al., 2014).

The rural digital divide stems from the high Capital Expenses (CAPEX) and low return of servicing the so-called last mile(s) for sparsely populated areas (Ovando et al., 2015). No surprisingly, 90% of the EU's rural population does not have access to Fiber-To-The-Premises (FTTP) or coaxial cable connections (European Comission, 2016). For rural households already covered by telephone lines, using modem dial-up and/or Digital Subscriber Lines (DSL) over phone copper wire is possible, however, due to the distances involved, speeds are limited and prices are not very attractive to the users (European Comission, 2016).

From a technical perspective, rural broadband can be achieved either by wireless infrastructure, such as satellite internet, WiMAX and LTE coverage (Ishmael et al., 2008), and/or wired infrastructure, such as FFTP, and coaxial cable (European Comission, 2016). Especially for the case of wired deployments, market forces alone are deemed ineffective to cover for such rural coverage gaps, and hence authorities have resorted to incentives for commercial operators. However, public policies are based upon universal coverage of geographic units and/or towns irrespective of final user adoption, which is generally smaller in rural areas due to digital illiteracy (Ovando et al., 2015). This further hinders the economic feasibility of such undertaking for commercial operators.

In particular, Vergara (2011) and Ovando (2015) have studied the case of rural Spain infrastructure deployments, with fixed and LTE infrastructures respectively. Relevant findings include that for the 25% of Spanish population, access to 30 Mbit/s broadband is not cost-effective by means of fixed infrastructure (FFTx, cable, or VSDL). LTE is more cost-effective in rural zones, but still fails to provide commercial returns for the last 1.5% of the population segment. In addition, other studies also deemed LTE to be commercially unviable for areas with less than 100 inhabitants per square kilometer (Feijoo Gonzalez and Gómez Barroso, 2013).

The issues being faced by commercial actors to provide rural broadband pave the way for WCNs emergence (Oliver et al., 2010), as a user-led, grass-roots alternative. Figure 64 depicts and illustrative rural WCN. Next we discuss the social and economic aspects of WCNs.

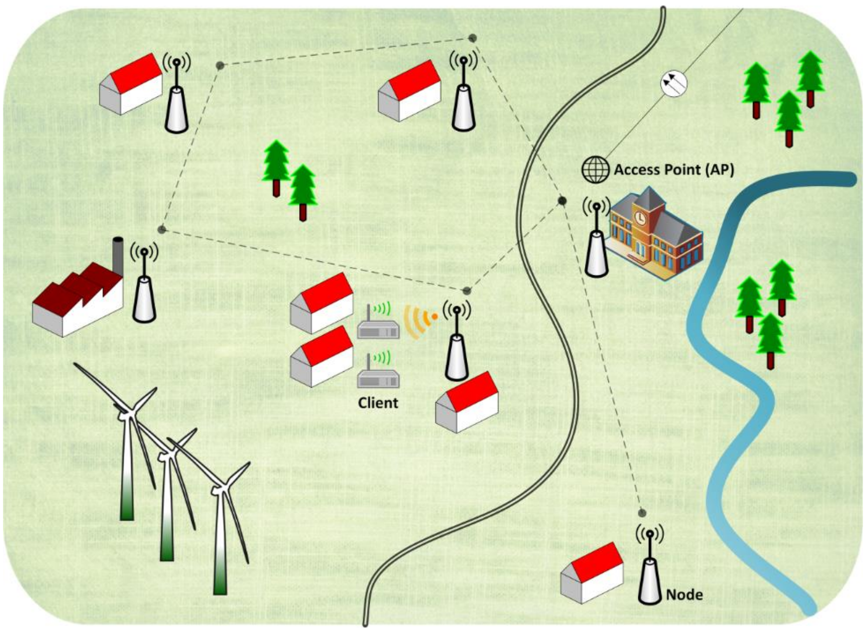

Figure 64. Notional Rural WCN based upon an Access point (AP) with high speed internet access. Several nodes connect wirelessly with the AP and relay to others. In the local area of each node, connectivity can be provided to pure clients with a short-range router.

8.1.2 Socio-economical aspects of a WCN

WCNs are usually non-for-profit arrangements (Abdelaal et al., 2009), while the provision of internet within WCNs can be subject to different cost-sharing, or for-profit arrangements. Notably, some commercial actors have attempted to commercialize similar services based on peer-to-peer signal piggyback (Abdelaal and Ali, 2007). WCNs are, on their early stages, based upon enthusiasts and technological-savvy users, who oftentimes see their implication in the WCN deployment as a hobby or social volunteering activity (Baig et al., 2015). This co-exists with alternative motivations of later stage adopters, who might weight-in cost-benefit considerations and compare WCN-based internet access with other solutions.

Abdelaal (2009) discusses the social capital driving WCNs and the forms of peer contribution enabling its emergence. Peers contribute to the global welfare with donations of money and hardware, by volunteering skills and time, and sharing their node connections as they are set up. Hence, besides a technical solution to a problem, WCNs are vehicles of digital empowerment and community building in countryside areas.

Motivations of ethical nature, and negative perception of commercial operators also play a role in WCN adoption (Oliver et al., 2010). User preferences for an *open* and *neutral* access to the net are commonplace in WCNs. *Open* refers to public disclosure of network protocols and structure, and *neutral* refers to lack of discrimination between traffic types or protocols (Baig et al., 2015). Such motivations are accentuated amongst peers of urban WCNs, where broadband is widely available and economic reasons apply less. Baig (2015) listed the benefits of WCNs motivating users. Amongst others, the author identifies *cost of access savings, citizens empowerment, universalization of access, and disappearance of infrastructure multiplicity* as benefits of WCNs.

8.1.3 WCN infrastructure and technological aspects

A survey by Avonts (2013) estimated about 63 Community Networks (CN) to be active worldwide and proceeded to analyze 19 of them. Most CNs are WCNs, while a small amount are based on fixed infrastructure. From the analyzed CNs, 53% use a subscription method to keep record of the active users, and the rest have little knowledge of the amount of subscribers in the network. WCNs are always grown around some form of social community organization which provides technical resources and advice to new users. More than half of the surveyed WCNs have created some form of legal entity, such as foundations or NGOs, to foster and protect their resources. In a few WCNs, the community organization also actively supervises the network topology, authorizing new node placements (21% of cases) and link establishment (25% of cases). In the rest of cases, new users are free to attach to the network, however they are encouraged to register their locations in the community database.

WCNs typically make use of Wi-Fi, WiMAX, or a blend of both technologies. WiMAX supports long-range, directional links up to 30 km depending on local orography (Ubiquiti Networks, 2017). LTE technologies are generally not attractive to WCNs due to regulatory issues, as they require obtaining spectrum licenses. WCNs make intensive use of COTS components and inexpensive consumer electronics. As reported by Avonts, 53% of WCNs use Ubiquiti products (Ubiquiti Networks, 2017) for their radio links and 20% use TP-link as main hardware vendor. Hence we can safely use their products' typical performances to characterize WCNs.

In terms of routing protocols, 53% of WCNs use OSLR, 16% use BGP, and 11% use the BATMAN protocol. For a brief review of network protocols, see section 4.5.1.1 .The analyzed CNs have an average of 67 links between nodes, with 3 of them exceeding 1,000 links, and the biggest one, Guifi.net, having more than 10,000 links.

8.1.4 The case of Guifi.net

The largest WCN community nowadays is Guifi.net, which has more than 30,000 nodes around the world, most of them localized in the region of Catalonia, eastern Spain. Figure 65 depicts the quick growth experimented by this WCN since its conception in 2004. Guifi.net is organized through a Wireless Commons charter agreement (Oliver et al., 2010) which defines the network as a shared resource, encourages sharing of knowledge between the users and declares the Industrial, Scientific and Medical (ISM) portion of the spectrum to be a common, public resource.

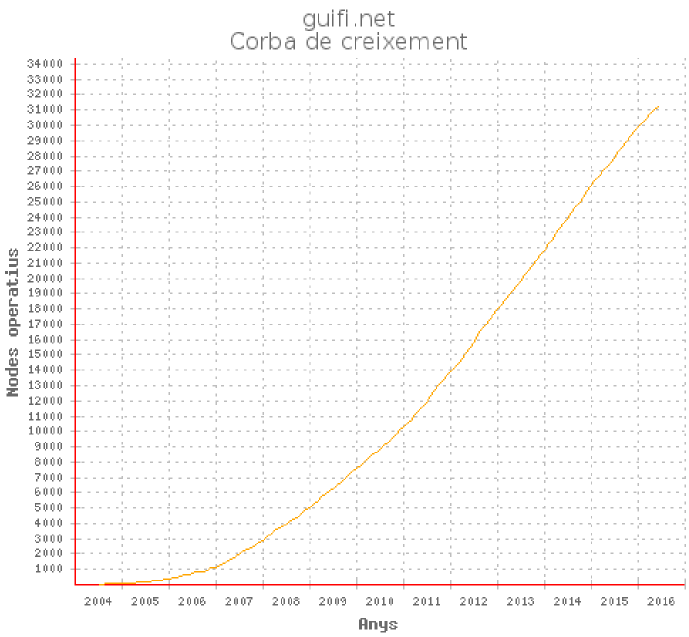

Figure 65. Growth of guifi.net accounted in operative nodes (vertical axis) between year 2004 and 2016 (horizontal axis), from (Fundació Privada per a la Xarxa Oberta, Lliure i Neutral Guifi.net, 2017).

Guifi.net started as a rural WCN but nowadays also covers urban population centers, spanning an area of about 10,000 km^2 (Fundació Privada per a la Xarxa Oberta, Lliure i Neutral Guifi.net, 2017). Guifi's network structure, due to its size and surface, is based upon two tiers of nodes, *supernodes* and *client nodes*. *Supernodes* act as the network backbone, establishing point-to-point, high-range links with other supernodes. Client Nodes are attached to their local super-node and forward the connection to other clients nearby with inexpensive outdoor router equipment (Chieng and Ting, 2011). Guifi.net is a mesh network, even though its geographical sparseness does not allow implementing a fully connected network.

Figure 66 illustrates the network topology, base-graph and core-graph, for the Catalonia region as per 2012. The core-graph is a subgraph of the network, which excludes all the nodes with degree one, that is, not connecting to more than 1 node. Those are typically client nodes. The core graph of Guifi.net includes 735 nodes of the total of 10,625 active in 2012. These are supernodes. We can distinguish below several clusters of geographically close, densely connected nodes, connecting to other clusters through only a few edges. Hence, we can easily identify a few single-failure points which would entirely disconnect some of the graph sections. These issues have been pointed as potential network vulnerability (Vega et al., 2012). Despite infrastructure deployment risks and challenges, Guifi.net is a very successful endeavor, with tens of thousands of daily users.

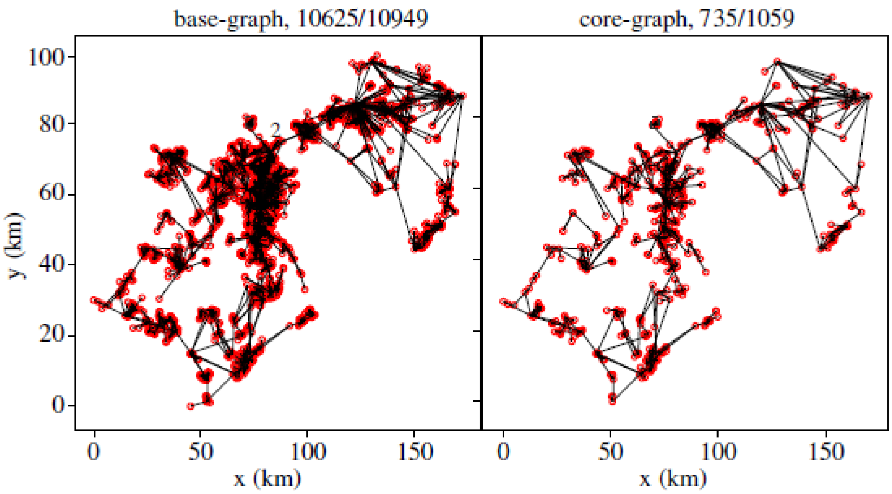

Figure 66. Base-graph containing all of the nodes for the Catalonia region and core-graph containing only the hub nodes, deleting leaf nodes and pure clients, from (Vega et al., 2012).

The case of Guifi.net illustrates the emergence and success of self-organized solutions to meet the latent needs of rural connectivity. In this case, these solutions spilled over a large territory and also tapped into urban population centers, creating a very large community-owned infrastructure. Many other WCNs of different sizes bridge the rural digital divide, based on peer-to-peer cooperation.

8.1.5 Characterization as a federation

WCNs like Guifi.net operate in patterns similar to FoS: the establishment and usefulness of the network depends on the voluntary cooperation of the systems and stakeholders involved. If we frame a new rural household internet connection as the system to be architected, we are faced with the decision if to become part of a WCN or to contract other services, when available. The value of the decision is subject to the evolution of the WCN and the number of adopters, as in all FoS. While in well-covered urban areas the economic tradeoff clearly favors the usage of existing infrastructure rather than deploying a wireless network node, in rural areas this tradeoff needs to be carefully analyzed. As in the case of satellites, peers in WCNs are deploying a communications infrastructure amongst themselves for their mutual benefit.

In early stages, regional governments and municipalities have supported the deployment of WCNs (Oliver et al., 2010), increasing its initial value and encouraging user adoption, hence showcasing a scenario of policy forcing. WCNs are a clear case, like most of FoS, of network externality issues, and potentially the *tragedy of the commons*, as some nodes piggyback internet connection from others. Besides economical motivations, WCNs respond also to social or political motivations which are more difficult to capture with the framework's utility-cost mindset. However, we expect the framework to be applicable in the rural cases, where WCNs are competitive with deploying infrastructure on cost grounds.

Table 33 maps the elements of an FoS to the specifics of WCNs introduced in this chapter. While coordination mechanisms are sometimes in place, WCNs do not explicitly manage the peers' locations or node capabilities.

Table 43. Application of FoS concepts to the WCN environment.

Conceptual element in FoS	**In WCNs**
Systems	Household internet connection
System manager/architect	House inhabitants
Resource/capability exchanged	Bandwidth
Architectural decisions	Internet access modalities

8.2 Applying the framework

Next we apply the framework for architecting federations to WCNs. The aim of this exercise, beyond illustrating from a new perspective the methodological aspects of the framework, is to study the deployment of WCNs bringing broadband connectivity to a rural area. The deployment is analyzed through the perspective of individual peer adoption, and highlighting its sensitivity to demographic, market and topological considerations.

8.2.1 Identify FoS context

We will analyze a generic rural stand-alone WCN, with a span of 100 km^2 and a low population density. Table 44 presents the rural population density across different countries. The upper limit depends largely on the local definitions of rurality, while the lower bound gives an estimation of the sparseness of the country's population.

Table 44. Typical rural population densities in households per square kilometer (HH/km^2). 3 dwellers per household have been assumed.

Country	Rural household density	Reference
Australia	0.2-4 HH/km2	(Riding et al., 2009)
Spain	1-30 HH/km2	(IDESCAT, 2016)
Russia	0.01-6 HH/km2	(Wikipedia, 2017)
US	0.6-64 HH/km	Us census bureau

Notwithstanding sensitivity assessments, we will focus here on the lower spectrum of population densities. These low densities are naturally more frequent in the areas without adequate LTE and DSL availability and hence of maximum appeal for WCNs as discussed in the first section 1 of this chapter. Note that household density does not map one to one to potential WCN user's density, as due its specific demographics and the digital divide itself, adoption is limited to a fraction to these households (Ovando et al., 2015). The density of potential WCN users –households interested in having broadband– might be half or less of the household density. The evaluation of alternatives requires of market information, ISP product bundle pricings and other ancillary parameters. In order to use a consistent set of data, we will focus on the rural Spain context when estimating the different parameters.

The household distribution on the territory will be generated by using two random uniform variables to obtain the Cartesian coordinates of each household. While densely populated areas can be arguably represented by a normal distribution, we deem best for generic, rural areas a uniform distribution. At {0,0} we place the AP. Figure 67 illustrates a random household map with 100 households over a 100 km2 territory.

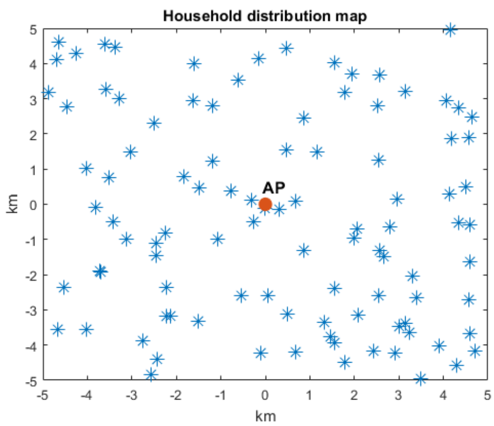

Figure 67. Example of uniformly distributed household set over a 100 km2 surface. The Access Point (AP) is located in the center of the region.

We will assume the network to have a unique Access Point (AP), source and sink of all traffic, with unlimited external bandwidth, as depicted in Figure 64. While large WCNs have many APs, and establish cross connections between them, small WCN usually grow from single AP. This AP represents a population center, or village, serviced by a conventional service provider using fiber optics. For the rest of the territory, we assume that no readily available options exist.

8.2.2 Formulate local utility and cost functions

The main reference performance for the user will be the speed, in Mbit/s, of their connection. We assume a symmetric uplink/downlink performance. The value of the bandwidth to the user can be proxied by segments, using the speed baskets identified by the EU in their broadband coverage reports (European Commission et al., 2015). For broadband, the basket thresholds are 1,12,30 and 100 Mbit/s.

We map here the performance to user utility with a piecewise linear function with 4 segments, considering a 0 utility to the user for speeds below 1 Mbit/s that are not broadband and utility 1 to speeds of 100 Mbit/s or more, which are at the limit of what is offered to private households in the market. For the first segments, we assume users are much more concerned about speed when their speed is in a lower segment, and hence half of the overall value is achieved at 12 Mbit/s. Figure 68 depicts the proposed performance-utility mapping. The sensitivity assessments in section 1.4 introduce different assumptions for customer value against performance.

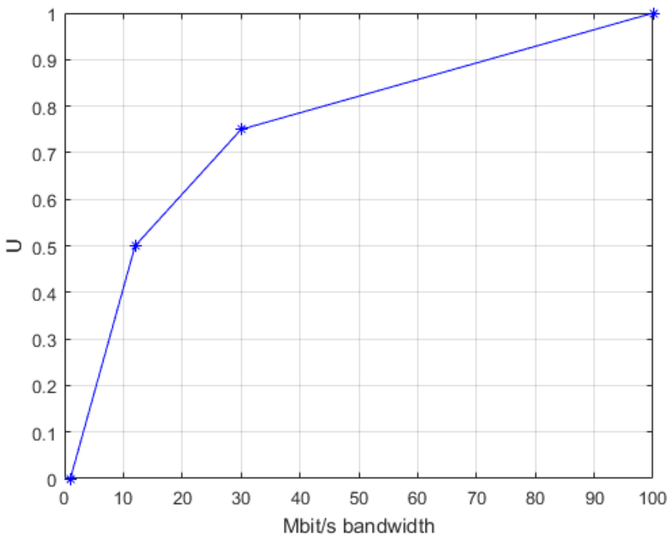

Figure 68. Bandwidth performance mapped to user utility for a residential connection.

The costs attached to the performances here will be referred to a monthly basis and include two main components: the amortization of the CAPEX expenses when applicable (fixed and WCN deployment options) and, as OPEX expenses, the cost of the necessary ISP subscription. In the case of Fixed and WCN options, this subscription is realized at the AP node.

The fixed infrastructure is assumed to be based on fiber optics, hence being FTTP. Other options are possible as Table 42 lists. Nonetheless, the expenses of deploying a fixed infrastructure are driven by the costs of civil works and labor and not by the costs of the actual wiring (Vergara Pardillo, 2011), hence we use the highest performance option. We will assume the fixed infrastructure is also a community effort and self-financed. However, due its high costs, we deem more realistic to model it as a centralized effort, probably directed by governmental institutions. Hence, we will compute the overall costs of deploying the FTTP to all households and distribute it amongst the peers uniformly, regardless their actual position. That is, citizens do not pay their personal cable deployment section by section, but subsidize through a coordinated effort the deployment of the whole infrastructure. Besides this CAPEX for FTTP deployment, each peer pays, at a commercial rate, OPEX expenses of contracting a commercial product that delivers internet connectivity through the FTTP and AP.

In the case of WCN, each peer invests as CAPEX the cost of the wireless equipment bringing connectivity to her household. This cost includes a pair of routers to establish the link on both edge sides. The WCN internet access spreads from the AP and hence the OPEX cost, as in the case of fixed infrastructure, stems from the contracting of commercial services at the AP. In addition, the WCN network edge capacity will constrain the amount of bandwidth needed at the source. The costs of maintenance for the infrastructures will not be modeled as they are smaller than the monthly amortizations of the CAPEX for both options.

Table 45 lists the data from the Spanish telecommunications market to estimate datarate performance and costs of WCN and FFTP, as a function of network edge length and edge capacities.

Table 45. Ancillary parameters needed to estimate final costs of FFTP and WCN connections. Cost values in USD 2016.

Parameter	Value	Reference
WCN CAPEX per edge	980 USD	(Fundació Privada per a la Xarxa Oberta, Lliure i Neutral Guifi.net, 2017)
FTTP capex	15,000 USD/km	(Vergara Pardillo, 2011)
WiMAX range	10-30 km	(Ubiquiti Networks, 2017)
WiMAX rate	150 Mbit/s	(Fundació Privada per a la Xarxa Oberta, Lliure i Neutral Guifi.net, 2017) (Ubiquiti Networks, 2017)
ISP rate per 100 Mbit/s, Spain	41.86 USD/month	(European Commission et al., 2015)
Infrastructure Amortization Period	20 yr	(Vergara Pardillo, 2011)

The CAPEX for FFTP varies widely depending on the orography and local labor costs. For Australia, (Townsend et al., 2013) values of 11,770 USD/km have been proposed. In the US, a thorough survey by an ISP provider proposed an empirical rules resulting in 11,526 USD/km. For Spain, Vergara (2011) developed a comprehensive cost model, from which we can extract approximate costs of 18,218 USD/km. Spain's values are more expensive due to the relatively more mountainous topography. We henceforth use a compromise value of 15,000 USD/km.

8.2.3 Enumerate architectural decisions

The decision for the peers involved in the WCN is what to use for their household internet access. The options we will consider are 2 products based on satellite connectivity, the option to use a WCN, and the option to deploy a fixed infrastructure. The cost of the last two options depends on the particular household distribution in the area of interest. Table 46 introduces the peer options, costs and performances associated.

Table 46. Internet connectivity options for a rural household. While we are focusing in the Spanish market, for consistency with the rest of case studies, we use USD 2016 as reference currency. To convert data from EUR, the EUR/USD exchange rate is used with the average of the applicable year, and then corrected for inflation.

Option	Costs associated	Performance	Reference
Satellite Internet 1	60.77 USD/month	12 Mbit/s	(Eurona sat, 2017)
Satellite Internet 2	106.4 USD/month	22 Mbit/s	(Eurona sat, 2017)
Fixed (FTTP)	Variable	100 Mbit/s	
WCN	Variable	Variable	

The first 3 options constitute the stand-alone tradespace of the each peer, and also belong to the SPF in most of the cases assessed in the results section.

8.2.4 Evaluate FoS configurations: compute synergy

Following the framework, we will compute the NBC of each peer and aggregate it as the overall FoS configuration synergy. For this, we require the costs and performance of the network. For satellite options those are fixed as per Table 46. For FFTP and WCN the costs depend on the amount of edges, and hence we need a network topology model. We use the Gastner-Newman (GN) network structure model.

The GN network structure is tailored to the growth prediction of spatially distributed networks (Gastner and Newman, 2006). This model has been widely used to model transportation networks, internet expansion and air traffic. We used it here to generate the allocation of edges and links. We make use of the Matlab™ Tools for Network Analysis developed at MIT strategic engineering research group (available at strategic.mit.edu) to generate NG networks.

8.2.4.1 Network modelling

A GN Network Structure model is adequate to simulate both to WCNs and FFTP edge allocations. A GN network spans all the nodes in the network, in a way such that they are connected to at least another node. GN networks have a central hub, from which the edges are built by minimizing the mean *effective* distances between nodes. The effective distance is a combination of the actual edge lengths and the amount of edges transverse from the hub to the periphery. The preference between both measures is controlled with a parameter *ß*. *ß=0* generates a classic shortest distance graph, while a *ß*=1 generates a direct connection from the hub to all the nodes. Figure 69 illustrates these different topologies.

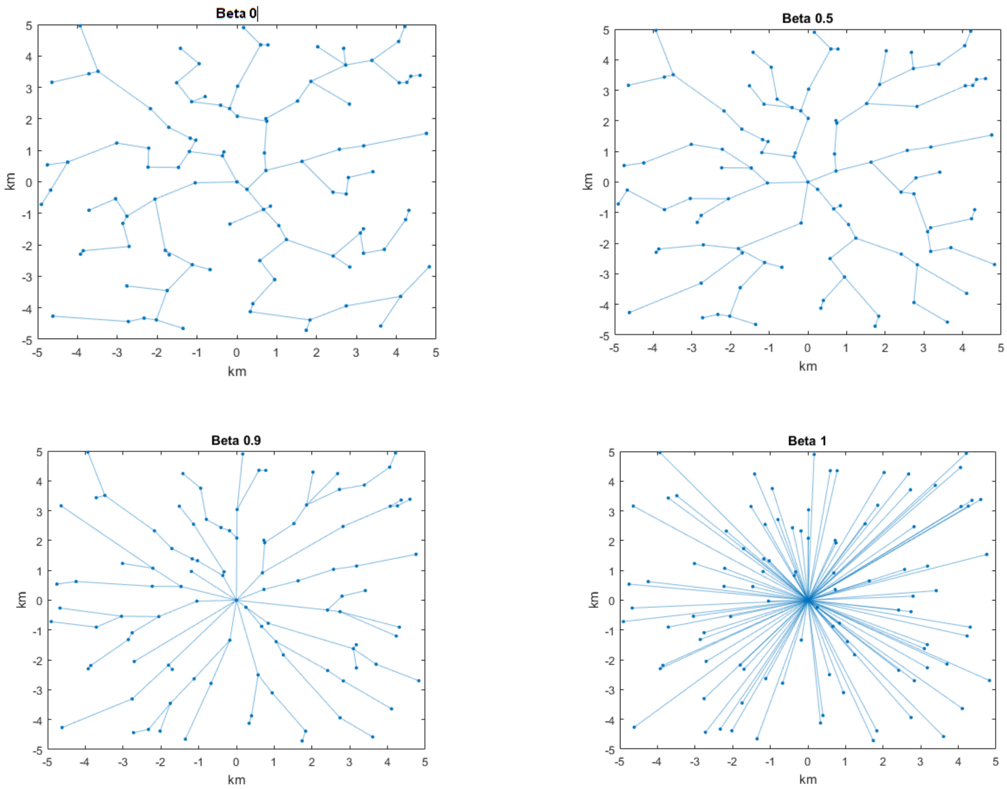

Figure 69. Different ß parameters generate different NG graphs over a set of nodes spatially distributed. The graph hub is located at {0,0}.

Note NG graphs and corresponding networks considered herewith are undirected. We call through this chapter *trunks* to the edges connecting the hub with other nodes, and *levels* to the set of nodes at a specific hop distance to the hub. That is, all the nodes that are exactly at a 2 edges distance from the hub (using the shortest path) constitute the second level of the network, and so on.

For the FTTP deployment, we will generate the edges with *ß*=0, as minimum distances are desirable for costs, and adding levels to the network does not affect the performance due to the large capacity of fiber optics trunk cables. However an intermediate value is desirable for WCN. We desire the edges to be short due to WiMAX range considerations, but, at the same time, the limited capacity of the trunks does affect the downstream datarate at subsequent network levels. Range allowing, increasing the number of direct hub-peer connections, or trunks, does improve the network capacity. We will start nominally with β=0.9 and add sensitivity assessments.

As for FFTP, the performance for each peer can be assumed to match the maximum user requirement: 100 Mbit/s. For WCNs, however, we generally need to solve a maximum network flow problem (Ford Jr and Fulkerson, 1956) taking into account the WiMAX datarate constraints. We will assume the same time-averaged allocation of bandwidth for all users, resulting from fair distributions of the capacity of the trunk they are connected to. Thanks to the network topology and this assumption, the problem becomes trivially the division of the trunk capacity to the amount of successor nodes.

8.2.4.2 Numerical model validation

Figure 70 illustrates a notional rural network posed for numerical validation of the model. The parameters needed to compute performance and cost for FTTP and WCN are found in Table 47.

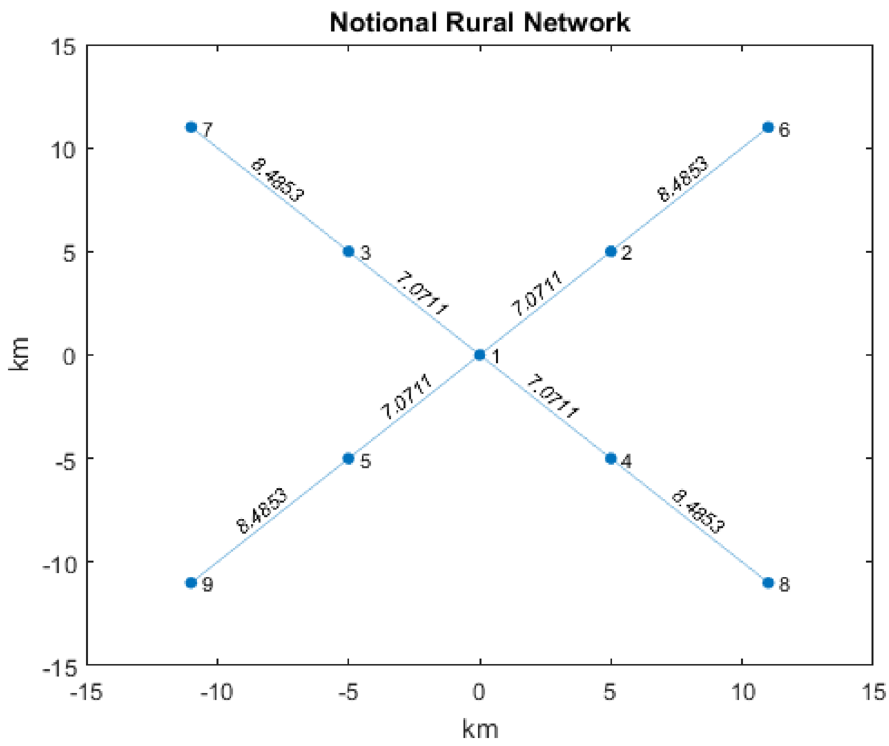

Figure 70. Notional Rural network with an AP and 9 nodes at locations {5,5}{-5,5}{5,-5}{-5,-5}{11,11}{-11,11},{11,-11}{-11,-11} km. Distances are shown at the edges.

The Synergy of the overall WCN and its distribution per peer follow from the aggregation of the local NBC as usual. Table 47 shows the results which agree with hand-made calculations in this case, verifying the model implementation.

Table 47. Numerical validation input parameters and output results, using notional data.

Input parameter	Value	Input parameter	Value
Households	8	Provider fare for 100 Mbit/s	50 USD
FFTP edge capacity	Inf	FO deployment CAPEX	10,000 USD/km
WiMAX edge capacity	450 Mbit/s	WiMAX deployment CAPEX	500 USD
Amortization period	20 yr	Max. user demand at node	100 Mbit/s
		Output comparison	
		Manual output	Implemented Model
Length of trunk edges		7.07 km	7.07 km
Length of second level edges		8.49 km	8.49 km
Monthly cost of WCN		416 USD	416 USD
Monthly cost FTTP		2993 USD	2993 USD
Synergy		2576 USD	2576 USD
Distributed Synergy		322 USD	322 USD

8.2.5 Apply policy to select FoS configurations

The next framework step is to navigate the MDP problem. Each peer has 3 options that represent opting out of the federation and taking options of her SPF, and an option to join the WCN. To enhance the generality and clarity of the results we will assume a deployment per levels instead as of individual peers. In this manner, at each FoS state the next level of users takes, as a block, the decision to join the WCN or not. Upon a negative, the next level cannot be served connectivity and the FoS configuration path terminates.

The actual household distribution is drawn from a pair of uniform random variables. Hence to be able to derive general insights, we will apply the Monte Carlo method. The next section estimates the necessary number of samples.

8.2.5.1 Monte Carlo evaluation

We now perform a set of sampling experiments on the statistical mean and standard deviation of the distributed synergy values of the last FoS state. Such experiments provide for convergence analysis. Said experiments use the same data as in the validation case, but with the demographics shown in Table 48.

Table 48. Demographics for the Monte Carlo experiments.

Parameter	Value
Network parameter beta	0.9
Households	100
Map side	10 (100km2)
Peer Density	1 HH/km2

The results oscillate only for a reduced amount of experiments based on 50-500 samples, before settling on a narrow, low frequency oscillation from approximately 1,000 samples. Figure 71 and Figure 72 show the evidence of convergence.

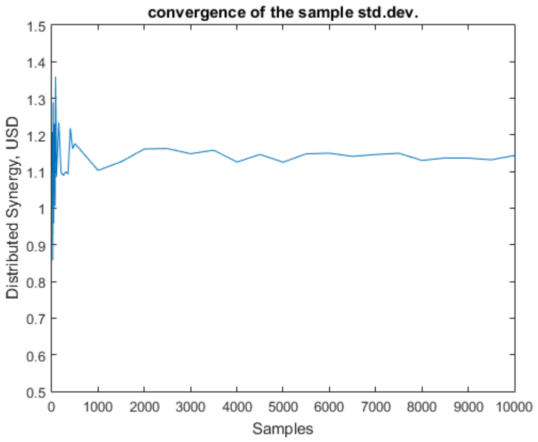

Figure 71. Standard deviation convergence for experiments up to 10,000 samples.

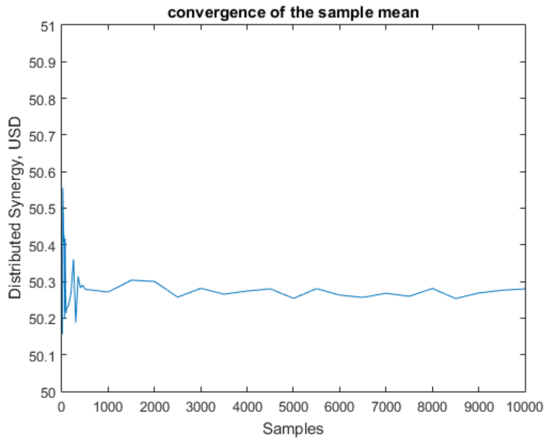

Figure 72. Mean convergence for experiments up to 10,000 samples.

Hence we pick 3000 samples as standard for our experiments. While the change of parameters could affect the necessary experiment size for convergence, this extra sampling should accommodate for any variations in the latter. Therefore, the results obtained in the next section all derive from 3,000 samples per case.

8.3 Results

This section describes the results of several framework runs with alternative parameters. The data presented in the plots includes the statistical information from the Monte Carlo sampling in the cases it does not hinder the clarity of the charts.

8.3.1 Nominal case

The first case run corresponds to the data in Table 45, Table 46, and Table 48, corresponding to a realistic case for rural Spain with 1 HH/km2 peer density. Results in synergy and distributed synergy amongst peers are presented by deployment levels, or FoS states. Figure 73 and Figure 74 show the results up to 12 deployment levels, which is the largest graph radius contained in the sampling set. Note the statistical quality of the data is reduced on the last two deployment levels due the naturally lesser occurrence of these cases in the Monte Carlo sampling. Nonetheless, we can observe clear trend for increase, saturation and decrease of synergy in this case study.

First focusing on the overall synergy, we can observe peers opt for federation in the WCN. However its overall synergy peaks about the 6^{th} deployment level, roughly corresponding to 60 households in this case. With this FoS configuration, we are offspringing 5 successor levels from each trunk. Considering the WiMAX datarate limit of 150 Mbit/s in this case, we are roughly providing 25 Mbit/s at each level. With one or two nodes per level, this is enough to deliver 50-75 % of the value to the peers. From this point on, adding more levels drops the overall performance and the mapped utility function switches to a steeper regime as Figure 68 shows.

Hence, the overall synergy stalls and only maintains a value of about 5,000 USD due the addition of new systems, even though the overall performance for individual peers is decreasing all through this federating process. If we instead look at the distributed synergy per peer, we can more dramatically see the aforementioned effects, as adding systems to the FoS does not soften the effect on distributed synergy value.

As Figure 74 shows, up to the 3rd level of deployment, the value of the federation is increasing for the individual peers. At the peak of distributed synergy, each peer is saving about 55 USD a month for their internet connection with respect to obtaining the same performance by other means, like satellite or fixed infrastructure. However, from that point on, forwarding the connection from the AP does hinder the distributed synergy as more *parasitic* nodes are being added.

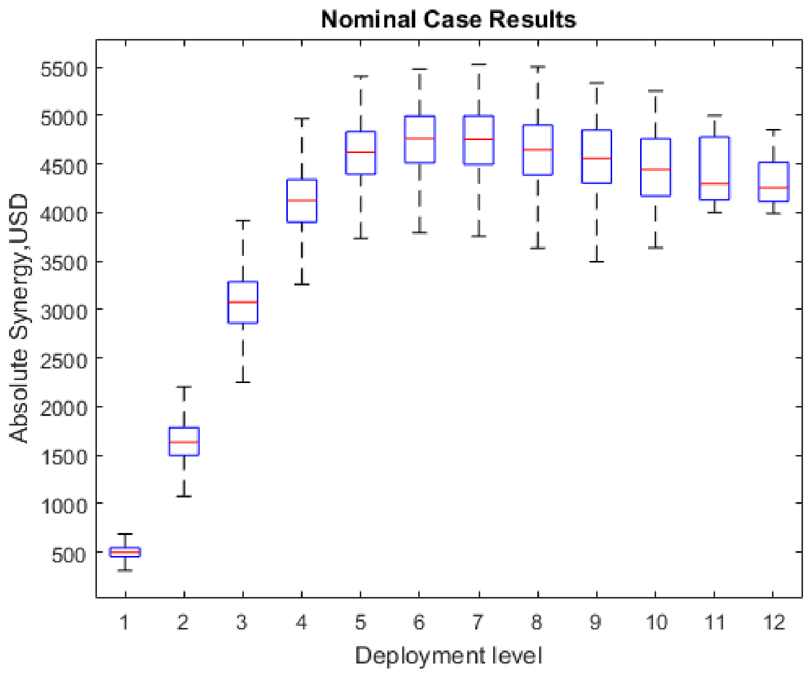

Figure 73. Total FoS Synergy through 12 deployment levels for the nominal case. Boxed data includes mean, 25 and 75 quartiles, and whiskers represent the range of the dataset including 99.3% of data coverage.

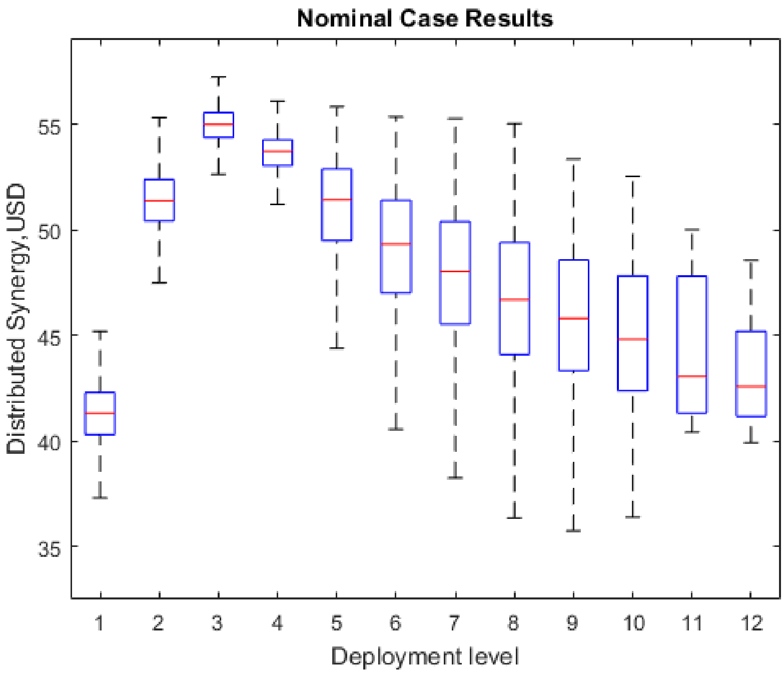

Figure 74. Distributed FoS Synergy through 12 deployment levels for the nominal case. Boxed data includes mean, 25 and 75 quartiles, and whiskers represent the range of the dataset including 99.3% of data coverage. Note the vertical range axis spans 30 to 60 USD for improved data visualization.

Note that while there is positive distributed synergy, there is no reason for any peer to drop from the FoS, as they still obtain better utility-cost than when using an architectural alternative. However, the perception of this dynamics can negatively affect the early adopters of the FoS who enjoyed better performances at FoS initial stages. This important feature was also distinguished in the FSS case and is reflected upon in this chapter's conclusions.

Note that in this case, the CAPEX and OPEX costs of FTTP are about 85-90 USD per peer and month, hence dominating the second satellite internet option and excluding the latter from the SPF. The costs per month and peer of WCN-based internet are typically about 20-40 USD, depending on the served bandwidth.

We have observed in this nominal results the influence of different utility function pieces, and WiMAX capabilities, to the stall and rate change of synergy through the FoS architectural path. This is explored in more detail in the following sensitivity assessment sections.

8.3.2 Sensitivity to demography

Now we study the FoS deployment with different amount of peers, maintaining and geographical area as 100 km2 and including more peers, from 10 to 500 households. This represents densities between 0.1 and 5 HH/km^2 acting as potential WCN peers. All other parameters are kept nominal. Figure 75 shows that, as expected, more densely populated areas suffer a synergy reduction as FFTP deployment starts being competitive. For very sparsely populated regions, FFTP is prohibitively expensive and hence the cost advantages per peer of adopting WCN are notorious. In such scenarios, the full network can be covered with just a few levels.

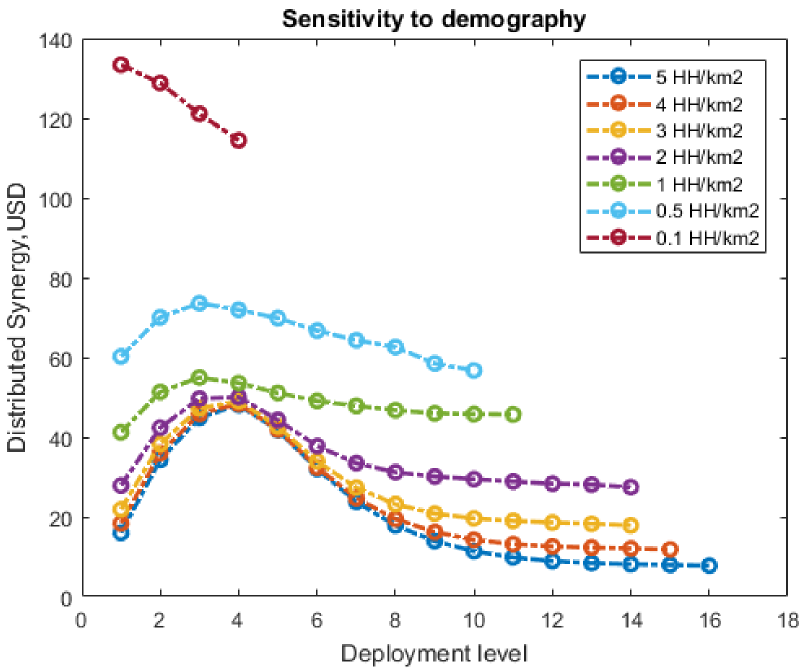

Figure 75. Sensitivity of distributed synergy per deployment level results to the area demography. Statistical data omitted for clarity: only mean values of the sample are represented. The difference in the amount of final levels stems from the application of the NG network structure.

8.3.3 Sensitivity to network topology

Next, we analyze the effect of topology through varying the NG network parameter *β*. We use ranges from 0.1 to 0.9. Recall lower *β* values generates more levels and less peers per trunk, while higher values minimize the amount of levels therefore attaching more nodes to each trunk. Figure 76 presents this sensitivity results. The initial peak in distributed synergy for a 3-level system is identical for all cases, while the following synergy decreases at a rate inversely proportional to *β*. The relative changes in distributed synergy are significant. In the final FoS configuration for *β=0.9* we obtain distributed synergy values of approximately 45 USD, while for the lowest *β* we roughly have cost advantages of 10 USD per month and peer. The lowest *β* topologies generate up to 17 levels in the WCN network.

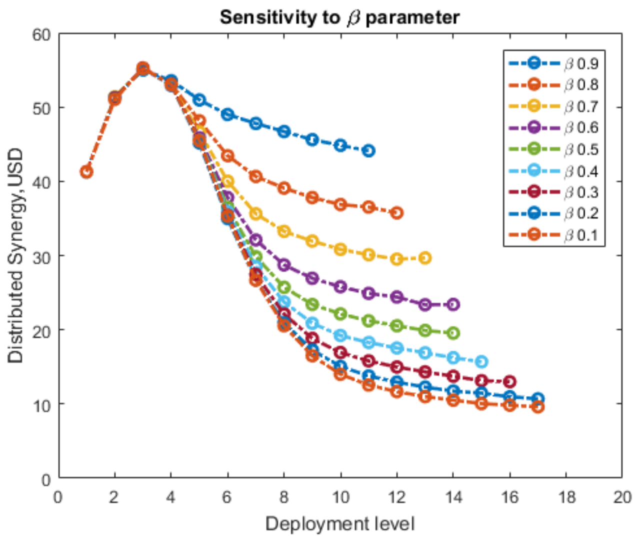

Figure 76. Sensitivity of distributed synergy per deployment level results to the beta parameter. Statistical data omitted for clarity: only mean values of the sample are represented.

8.3.4 Sensitivity to WiMAX data rate

WiMAX standards and products have evolved since its appearance, offering speeds from 30 Mbit/s to more than 400 Mbit/s for point to point links (Chieng and Ting, 2011; Ubiquiti Networks, 2017). The distances involved will also greatly affect the achievable rates. We now assess the effect of the edge bandwidth capacity in the attainable synergy. Figure 77 shows the FoS distributed synergy evolution in scenario from 30 Mbit/s to 450 Mbit/s.

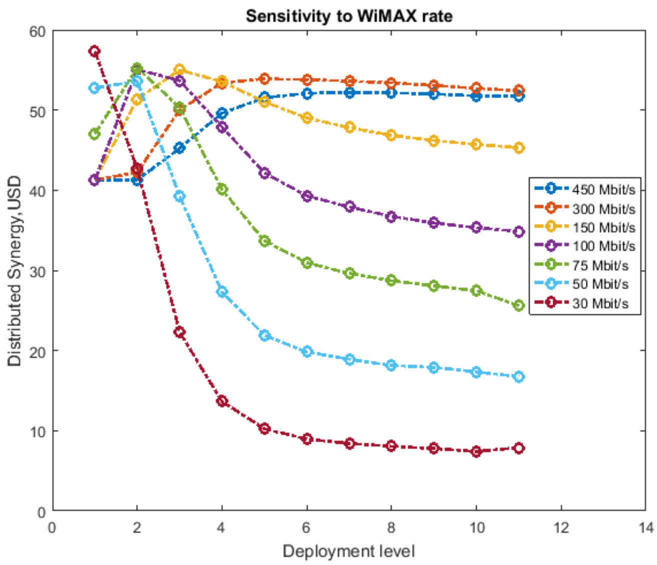

Figure 77. Sensitivity of the distributed synergy per level to the WiMAX darate. Statistical data omitted for clarity: only mean values of the sample are represented. Data of the last levels (>9) is relatively more noisy, as large graph radius scenarios are not often occurring in the sample.

Interestingly, for 30 Mbit/s datarate synergy continuously drops as peers join the FoS. For intermediate values of WiMAX rate, a peak and a subsequent drop distributed synergy appears in levels 2 or 3. Finally, for rates of 300 Mbit/s and above, the drop on synergy is very moderate and a high levels of distributed synergy can be maintained. As expected, larger capacity edges networks can overcome the effects of additional levels and branching factors in the network.

8.3.5 Sensitivity to utility functions

Now we analyze the effects of 1) changing the ranges of the piece-wise utility functions, and 2) using step functions instead of linear functions. The latter experiment aims to capture situations where peers are not sensitive to small performance changes and instead think about performance in plateaus. This matches the product structure offered by ISPs and the baskets considered in market studies (European Commission et al., 2015), and hence might more realistically represents internet bandwidth user value. Figure 78 illustrates the alternative utility-performance mapping functions.

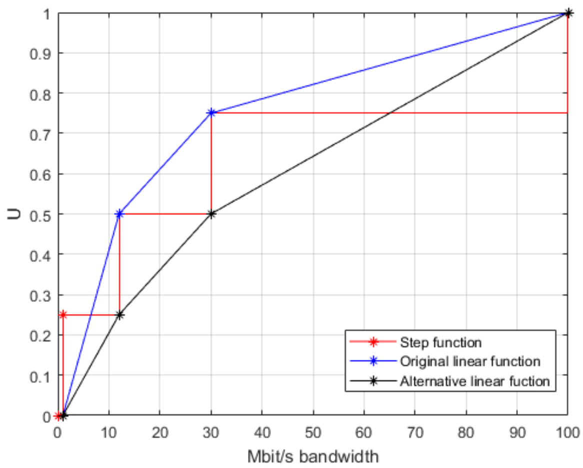

Figure 78. Alternative utility functions compared to the nominal assumption.

For the piece-wise linear alternative function, the results do not significantly change as confirmed by the distributed synergy chart shown in Figure 79.

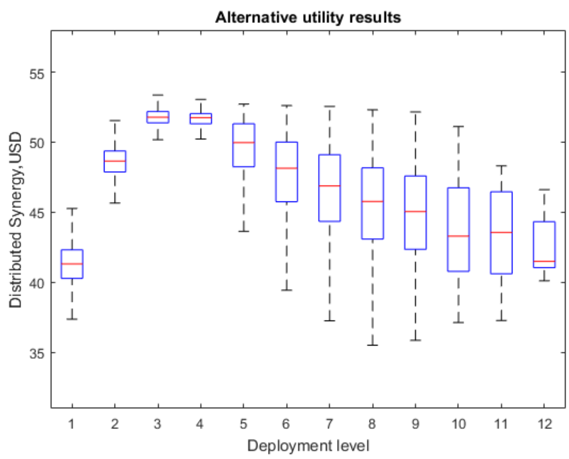

Figure 79. Distributed FoS Synergy through 12 deployment levels with the alternative piece-wise linear utility function.

For the step-function, the effect of the 30-100 Mbit range plateau is to maintain the distributed synergy at a value regardless of the addition of nodes, as Figure 80 shows.

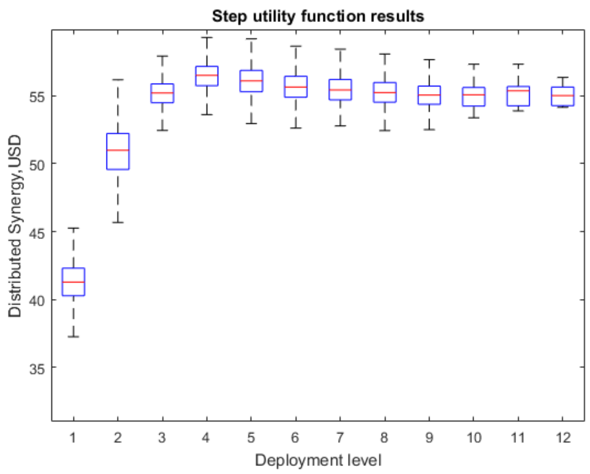

Figure 80. Distributed FoS Synergy through 12 deployment levels with the the step utility function.

8.3.6 Terminating the federation: aiming for negative synergy

Finally, we analyze the conditions for federation termination in WCNs. In this corner case, we analyze how a FoS starting with positive synergy can reach negative synergy condition and hence start losing peers. For all of the results presented so far, distributed synergy reached a positive plateau value. What we want to find here is a combination of low β, high population density and low WiMAX rate, such that negative synergy is achieved. For illustrative purposes, we force a positive adoption policy to highlight negative synergy at final stages of deployment, even though the conventional application of the framework would terminate upon reaching synergy 0. A preliminary exploration identified, for example, a 2 HH/km^2, β=0.5 and a WiMAX rate of 50 Mbit/s to lead to negative synergy from the 9th deployment level (i.e., the sample average is below 0). After exploring the parameter combinatorial we conclude that WCNs can eventually fail on later stages, in the parameter space Figure 81 depicts.

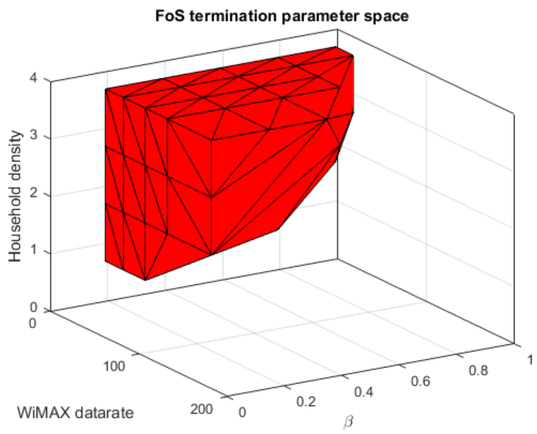

Figure 81. Parameter space for which, after a few levels, the FoS tends to negative synergy and hence the WCN could be discontinued.

8.4 Conclusions

This chapter's analysis of WCNs highlights their potential to bridge the rural digital divide, a potential which is slowly being realized worldwide. Remarkably, the results in this section show that WCNs are typically appealing for the first levels of peers in the network, those closer to the AP. As the WCN extends, the advantages for each individual peer decrease unless new access points are made available.

We have shown in a Spanish rural setting, with 2-5 HH/km^2 density, each peer can save about 10-20 USD per month, and values closer to 50 and 60 USD when density is 1 HH/km^2 or less. This cost advantages are respect to satellite and fixed infrastructure alternatives. Such benefits are sustainable up to 8-12 network levels if the wireless links capacity matches the demand.

For higher population density situations, the performance of fixed infrastructure becomes competitive with the wireless. While deploying a fixed infrastructure can also be subject to a community effort, the wireless case is more compelling for our purposes. The application of the framework rendered both case-dependent and general insights.

8.4.1 Case study insights

Table 49 summarizes the effects influencing the emergence of WCNs as a federation, and its potential demise.

Table 49. Effects of different parameters on the prospects for WCN federation.

Parameter	Effect on benefits	Example values run	Recommendations for federation emergence
Wireless edge capacity	↑↑	[30:450] Mbit/s	Best kept above 30 Mbit/s.
Network topology - levels	↓	[0.1:0.1:0.9] β	Minimize the amount of levels as much as possible, while respecting maximum ranges for edges.
Peer density	↓	From 0.1 to 5 HH/km^2	The less peer density the better, as fixed infrastructure becomes competitive.
Utility function	-	Nominal shape, alternative and step function	Users simply need be positively sensitive to the bandwidth performance

Some combinations of the parameters proposed can lead to the termination of the federation as section 8.3.6 shows. Remarkably, for all combinations of parameters, the distributed synergy peeks in the federation at about the 3rd deployment level. This is a combined result of the maximum user demand, set at 100 Mbit/s, and the ranges of WiMAX capacities available. While the peaking level is a specific result of this case study, the trend for synergy to increase, peak and stall has been identified in the other case studies.

With a 450 Mbit/s capacity on the edge, three network levels can be served up to 100 Mbit/s. Assuming a single peer per level, such peer would achieve utility 1 in their local tradespace. This is no longer possible at the 4th level. On the other hand, at the minimum edge capacity of 30 Mbit/s, about 10 Mbit/s are served to the 3 first levels. This brings them close to the first performance basket, 12 Mbit/s. From that point on, the descent in utility is steeper and this effects the distributed synergy at the 4th level. Hence we can conclude that the optimal distribution of high bandwidth APs in a WCN is one per each node cluster up to 3 edges in radii.

8.4.2 Theoretical insights

Distributed Synergy appears to be a saturating property of federations, eventually constrained by the connectivity between the peers and the overall federation contact to the infrastructure. In the case of FSS, the restrictions were embodied by space networks and connections to the ground stations, and in WCN case, the AP location and edge capacity. For WCN the network topology is fixed, which makes more obvious the issues of adding additional network layers without close contact to the infrastructure. This issues also affected FSS, but the topology dynamics made them less apparent. Nonetheless, in the ride-sourcing study, synergy saturation was also observable, but attributable to the market limits, not boundaries of technological nature.

When new peers do not contribute with additional resources –direct AP connectivity in this case, or GS contacts in FSS– the distributed synergy suffers, while the overall can keep increasing as new systems are served. This is indeed an example of the *tragedy of the commons.* This poses a new question: could existing FoS peers, or systems, refuse to cooperate with new adopters, even in positive synergy scenarios, shall they perceive a decrease in distributed synergy?

In first instance, this is driven by the openness of the federation, its technical standards and communication protocols, which might in some cases not make possible to easily select with which peers to communicate or not. In the case of FSS and WCN, it would seem technically possible to selectively refuse to cooperate with individual peers. In the case of WCN it could be more complicated due to the ethics behind real-world WCNs, but not impossible. In the case of ride-sourcing, the federation is open and there are no easy ways to close it to riders. Nonetheless, an unbalanced excess of drivers or riders can lead to its collapse, and the role of the coordination platform is to avoid these situations by using the pricing controls. Similarly, in WCNs and FSS, we desire to achieve balance between resources demand and offer.

Additionally, there are two reasons why existing FoS peers would accept to cooperate with non-contributing peers. First of all, we need to recall the discussions of distributed synergy and incentives. While distributed synergy is a peer average of the global cost advantages' present, incentives can change the local value of federating for each peer. Positive synergy, as described in the approach chapter, implies opportunity for the exchange of incentives; in the WCN case, peripheral nodes could reward their peers providing bridges to the AP. This would compensate the latter for the potential synergy decrease, and still make WCN an appealing option for the former.

The second reason for accepting new peers is expanding the network into potential resource pools. That means that, by adding nodes, a FoS improves its chances to connect at later stages to another AP, or a new satellite and its additional GS infrastructure. The value of such scenarios is bounded by a discount factor, as analyzed in the first case study.

Eventually, synergy can reach below 0 scenarios, where all peers shall stop cooperating and the FoS could face its dismantling. This chapter's study allowed us to characterize a FoS termination. In a WCN, dense demography, coupled with increased network levels and low edge capacity, can lead to unsustainable FoS situations. In such circumstances, unbounded FoS growth leads to the reduction of individual peer value up to the point that federating is not Pareto-optimal in any of the participant's tradespace. Due to user lock-in, we do not expect the FoS to immediately terminate in such a situation, and it could be argued that the last peers in would be the first to drop, returning the FoS to a sustainable state. However, it is desirable that such natural regulation is overseen and guided by a coordination mechanism, to keep FoS benefits well above a marginal state.

Chapter 9 Conclusions

This thesis has developed and applied a framework for architecting FoS. Being the first effort in the literature of this type, this work covers FoS definition and identification aspects, in addition to quantitative analysis of federation benefits.

Federations of Systems are a type of SoS. Since its advent as a formal academic topic in 1998, SoS research has focused on definitions and disquisitions of qualitative nature, and has almost entirely been devoted to directed SoS, as we review in chapter 4 of this thesis. This work instead delves into FoS, which fall under the definitions of *virtual SoS*, a cooperation agreement between systems for mutual benefit and without explicit overarching goals.

The motivation to explore FoS originated in satellite federations. This thesis tackles a generalized problem of architecting FoS, suitable to satellite federations but also to other types of FoS. Such architecting problem is subject to design coupling between systems, and the uncertainties over the federation evolution. To structure and guide this thesis' work on such problems, two specific research questions were posed. Figure 82 illustrates the thesis' flow from problem to solution.

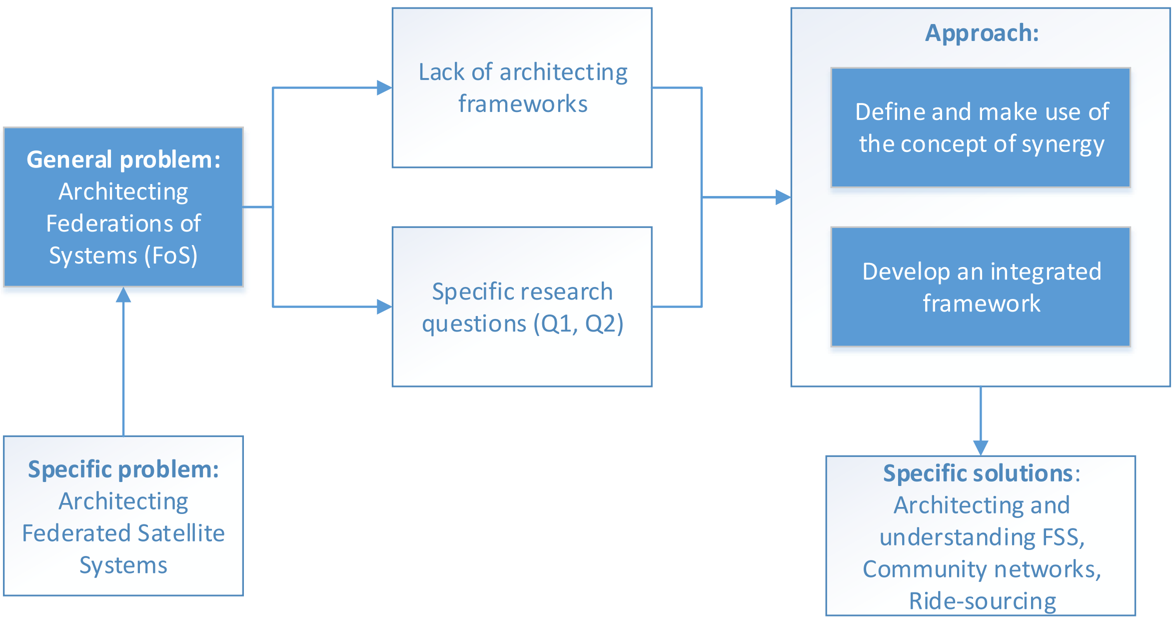

Figure 82. Conceptual vision of the thesis.

To answer the research questions, we have analyzed FoS with a tailored approach, using common components of the systems' architect toolkit, such as tradespace exploration, Pareto dominance and utility functions, and also the more general tool of MDP. Last but not least, we have defined mathematically and made use of a specific theoretical construct, ***synergy***, relatively unused in systems engineering but with parallels in economy, to support our dissertation about FoS.

These components configure the proposed framework, with which we analyzed 3 case studies concerning satellite EO, WCN and ridesourcing. First, we summarize here the findings of said case studies.

9.1 Case study findings

We briefly summarize here the key points from the different case studies, which are available in their respective conclusions sections. In our second case study, ridesourcing, we analyzed the NYC taxi transportation market and the emergence of Uber with a focus on retrospective validation. The case has verified the framework's approach, and matched the available data. Moreover, it yielded ridesourcing insights; in the NYC market about 10,000 riders/h and 4,000 active drivers are needed to reach self-sustaining operations, beneficial to both parties. In a large location as NYC, an upfront investment in driver and rider incentives of about 290 MUSD is required before the FoS is operational and distributing benefits. Therefore, this is only within the reach of capital-intensive ventures. Moreover, we have confirmed the notion, already proposed by many authors, that the difference in operating costs of Uber vehicles and licenced taxicabs is at the center of the former's advantage.

In the case of WCNs, we confirmed the advantages they hold in rural areas where the population density is below 5 HH/km^2. For each household, the savings of establishing a WCN with respects to wired infrastructure or satellite internet are about 10-60 USD monthly. For these reasons, and due the existence of suitable WiMAX technologies, several community wireless networks have been already established in the world. We have also identified the challenges they face. The tragedy of the commons is a typical failure mode of WCNs and hereby we recommend, with the technologies currently available, to deploy such networks in a way that households are not further than 3 edges from a high bandwidth AP.

For FSS we showed that just 2 satellites are enough to achieve a beneficial, early stage federation. This result is relevant, as it shows very early returns of federating, and encourages practitioners to implement federation testbeds. The FSS case was focused at LEO EO missions, and looked at a parameter space including interface costs, discount rates, architectural cost constraints, cooperation constraints, and sensitivity to different stakeholder utility functions.

From all the indicators of federation benefits used in chapter 6, the most intuitive is to use a % of the missions' communication and ground segment costs. Depending on all the parameters listed above, it was shown that the benefits of FSS can amount from 10% to 40% of the mission communications subsystem budget, including ground stations, satellite hardware, and operations. Moreover, using a federated approach to downlink mission data beats in cost the usage of geostationary relays.

FSS has non-negligible interface costs, which drive the achievable benefits. In chapter 6 we have assessed 24 possible price tags for the lifetime cost of the FSS interface and concluded it shall be kept below 15 or if possible, 10 MUSD. Now we briefly explore, for the sake of completion, the potential costs of such an interface.

9.1.1 Note on FSS costs and interface implementation

Is it possible to deploy FSS interfaces with a total lifecycle cost below 10 MUSD? Ballpark estimates of ISL and other components seem to show it feasible. Ongoing industrial R&D on FSS (ONION consortium, 2016) predicts the FSS interface for LEO to weight about 20 kg. The development costs of manufacturing a radio payload of 20 kilograms can be estimated using the USCM8 model (Wertz et al., 2011) to be roughly 14 MUSD in 2016. The recurrent manufacturing costs, with the same model, would be about 4.23 MUSD. If we conservatively spread the development costs across 10 units, without taking advantage of any economies of scale, we can establish a cost per unit of roughly 5.6 MUSD.

The impacts of this payload on the spacecraft might be on the order of an additional 3-6 kg in terms of additional power generation (Lluch i Cruz and Golkar, 2014), totaling 26 kg added to the spacecraft. The launch cost of adding 26 kg to the spacecraft can also be estimated. Launch costs widely vary depending on the launcher, being typically between 5,000 and 15,000 USD/kg (Wertz et al., 2011). Using a figure of 10,000 USD/kg, we obtain an additional launch cost of 0.26 MUSD.

Finally, we shall estimate the operation costs and the changes on the ground data dissemination infrastructure needed. The additional operational needs are 1) to maintain the ISL contacts schedule and coordination between missions, and 2) to separate the ground the data retrieved from every spacecraft and send it to its actual owner using terrestrial infrastructure. Since the space-to-ground throughput of missions is unchanged, arguably most of the ground segment hardware capabilities, including terrestrial bandwidth from the ground station to the payload data centers, can remain the same. The data separation can be automated and have a limited cost burden. However the first issue, the coordination and scheduling of the ISL contacts might require of a specialized member on the spacecraft operations team. Adding such a member could cost about 2 MUSD through 10 years lifetime (Wertz et al., 2011). Hence, the total lifecycle cost of the FSS interface can be preliminary estimated to be 7.86 MUSD per spacecraft.

In the nominal case study for FSS, such interface expenses would yield average cost savings for missions of about 17.26 MUSD (Chapter 6). Hence, under realistic assumptions, FSS holds the promise to improve mission's cost effectiveness.

We now continue the discussion shifting into theoretical findings. Let us start the discussion here with a reminder of the research questions and how this work's structure and content addresses them.

9.2 Research questions and theoretical findings

The first research question posed was *'How can we measure synergy between a set of engineering systems?'* Which decomposed in '*what is synergy'*, and '*under what conditions does it appear'*. This question drove the development of this work's approach.

9.2.1 How can we measure synergy?

We defined synergy formally as the aggregation of the Net Benefit of Cooperation (NBC) of the systems participating in a federation. The NBC is the benefit experienced by each of the federation members when cooperating, compared to its individual capabilities and after subtracting any additional costs incurred by federating, such as interface adoption and incentives for other systems. Hence a positive NBC for a system means it is better off cooperating that operating in isolation.

The NBC and therefore the synergy for each system was measured by evaluation of the local utility and cost of federating for each system, and its comparison against the SPF, the Pareto front of the system when performing an isolated tradespace exploration of its potential architectures.

The issue of consolidating the units of NBC across systems and hence being able to assess the advantages of each particular federation configuration has been given much attention in this work. The preparation work to this thesis and some of the works of the literature for SoS do use global measures of utility, external to each of the systems present, to measure the merits of a given FoS configuration. This is not justifiable theoretically due the absence of overarching goals for cooperating and global stakeholders, as we have argued in this thesis. If there is no encompassing motive for the cooperation rather than the benefit of the systems themselves, it is not easy to justify the existence of global utility functions to assess the FoS.

This thesis' approach avoids this shortfall by consolidating NBCs in units with similar meaning and value amongst all system. We used the economic returns of the architectures to assess NBC and synergy. This was simple enough to apply in the ridesourcing case study. However it generated new challenges with systems like EO satellites, whose operations have very broad and diffuse economic returns. In that case, we used the system Standalone Pareto Front, SPF, as a map between the stakeholder utility of the system and its economic value. The cost of a Pareto-dominant system architecture does not give an absolute idea of its merit, but due the nature of Pareto-dominance, it does give a monotonic, well-behaved valuation of a certain set of system's capabilities. Hence we can use this to compare merits of architectures, and we do in the satellite systems and WCN case study. When using this technique, the NBC can be interpreted as "the cost advantage obtained of federating by a system, with respect to obtaining the same set of performances on its own".

Mathematically, we defined synergy as a linear aggregation of NBCs. A sum is the simplest and most intuitive way to aggregate NBCs. A sum is convenient also when reasoning about incentives and how 'far' is a given FoS configuration from being beneficial for the parties involved. However, if we had defined synergy as a multiplication or any other function, the fundamentals of our approach would still hold. The defining point here is how we compute the NBCs. The NBCs of participant systems are oftentimes obtained via highly non-linear models –as detailed in the 3 case studies–. Hence, what really captures the benefits of cooperating and associated non-linearity are the performance models.

Table 50. The origin of synergy.

Origin of synergy	**Implementation mechanism**	**Examples from case studies**	**Types of synergy**
Functional emergence	Generation of a new function, not existing before by federating.	Improved latency by relaying in FSS.	Strong synergy, *superadditive.*
Resource reallocation	Shared access to an infrastructure or capability, shared amortization of an investment	Bandwidth sharing, peer-to-peer car transport.	Weak synergy, *subadditive.*

We have highlighted in this case studies several applied mechanisms for synergy emergence. Observing the case studies, we can connect the appearance of synergy to 1) classical functional emergence, and 2) the re-allocation of resources such that a more cost-effective configuration is achieved. Table 50 shows our findings in this aspect and how it connects to the themes of strong and weak synergy defined in the approach.

The synergy created, or the value of the federation, naturally depends on the costs of the adoption interface, the utility perceptions and performance needs of the systems' involved, and very importantly, on the topology of the FoS as a network and the access to the resources by peers. In FSS and WCN, a very important condition for the appearance of synergy is the difficulty to access a specific infrastructure and/or resource, which justifies the networking. Analogously, in the ridesourcing context, the scarce resource or infrastructure are the taxi medallions or licenses. In this case, however, the ridesourcing federation did not provide access to them, but made them irrelevant for the operation of the FoS.

9.2.2 How can we predict the formation and evolution of a federation?

The second research question posed was '*How can we predict the formation and evolution of a federation of engineering systems?*' And included: *'What are the effects of federating on a particular system?'* and '*How can we influence the long-term evolution of a federation?*'

The first branch of the question, about the effects of federating on each particular system, is integrally connected to our definition of synergy, expressed by each of the local system's NBCs. The proposed approach captures the effects of federating in a particular system through utility and cost displacements in its local tradespace. This is illustrated in detail in the results section for FSS, section 6.2.2.

The formation, evolution and temporal aspects of this question fostered the inclusion of a dynamic aspect in the framework. To model the emergence, progression and potential collapse of federations we analyzed the FoS deployment process, assessing the design of a system (or a batch of systems) one at a time. The decisions of each architect regarding the design of their system were modelled as actions in an MDP, which expanded as more systems were added. The decision space of each architect considered the federation upstream and downstream influences, yet the authority to federate or not remained local.

Depending on the case, we evaluated individual systems, or batches when the amount of systems or peers to consider was large. The framework predicted the emergence of 3 federations. The motivation of this work, FSS, has not been implemented yet and we here demonstrated the conditions favourable to its emergence. WCNs do exist nowadays and our general analysis of them identifies their value, but also the challenges they face. Finally, we retrospectively predicted the emergence and evolution of ridesourcing services in the applied case of Uber in NYC. The emergence modes of a federation are classified here in *no emergence, emergence with initial negative states, and spontaneous emergence.*

No emergence concerns the cases where, simply due the high costs of FoS interface, or due lack of needs for additional performance from the system's stakeholders, federating is not beneficial to the potential peers, regardless of the amount of peers present. This is a common occurrence for many parameter combinations in FSS and ridesourcing. The framework process is aimed at detecting highest synergy FoS paths, and it ruled out hundreds and thousands of negative results.

In other cases, *emergence requires of initial negative synergy states*. This is the case of the federations in FSS and resourcing, which require of incurring in losses or initial sub-optimal design decisions until synergy can be achieved. Such early adoption problem can be offset either by 1) external incentives, or 2) being able to capture future, discounted benefits in the local tradespace. The first option applied to ridesourcing, and the second we exemplified in the FSS case.

Emergence is spontaneous in cases like WCN. This means that from the first peers it is beneficial to federate. This can occur if the SPF features very expensive options and the federated interface delivers value independently of the existence of other peers. For the first peer of WCN, that connects directly to an AP wirelessly, the FoS interface delivers value, even if additional peers do not connect through. This leaves us the interesting idea that the challenges for FoS adoption can be mitigated if the FoS interface can bring value before other peers exists, for instance through alternative usage modes.

As per the collapse of federations, we used the WCN case, and ridesourcing to a lesser extent, to showcase different failure modes of a federation. We distinguish here two *failure modes* of a federation, and we call them *tragedy of the commons*, and *congestion*.

The failure mode of *tragedy of the commons*, as related in the case studies, stems from the addition of systems which do not contribute with resources to the FoS, but only make use of them. Even if they are ready to incentivize, that is, pay, for accessing the resources of their peers, the synergy of the FoS eventually decreases as the performance for all systems decays. Eventually, the performance and cost experienced by the peers could fall below the threshold of their SPF and hence the motivation to cooperate vanish.

In the case of ridesourcing, we observed *congestion* failure modes. Unbalanced numbers of drivers and riders could make the FoS unusable for the peers. This was termed in the case study 'game over situations'. This was especially critical in ridesourcing as the switching costs for riders are very little –they can just hail a conventional cab–. The advantages of federating in ridesourcing are more marginal and have less peer lock-in than in FSS and WCN. While not explicitly studied on the latter cases, we can envision the same type of issues in FSS and WCN networks. For instance, placing many requests for bandwidth to the network in both cases could lead to congestion and this type of failure mode.

The failure modes can be kept in check by the coordination mechanisms of the FoS, and hence are recoverable with proper governance processes in place. The actions of individual peers also influence the dynamics of the federation.

9.2.3 About control and influence dynamics

In the proposed approach, the instruments to control and influence the evolution and outcomes of the federation are allocated formally in the MDP action nodes. In the case of FSS, the MDP actions correspond to the architect design decisions. Hence every architect can influence the federation evolution. A clear example of each of the systems' leverage in FSS is their ability to choose amongst different federated options, all beneficial but to different degrees. Architects within the FSS can choose to deploy costly infrastructure and share their access, hence benefiting in the long term a larger amount of federates, or conversely go for more modest solutions which benefit all peers but to a lesser degree. We illustrated these options with the usage of different cost constraints in the FSS case study, emulating different behaviors of the architects. The adoption of more expensive local system architectures led to additional synergy in an approximately linear relationship.

In the case of WCN, the decision of each batch of federates, organized in network levels, could effectively preclude future systems from federating. Due the distance involved, households further from the access point depend on their peers to forward their connections.

Hence, the differences in network topology between these cases impact the influence and control exerted by the peers. Our WCN exhibited a network topology of extended-star type, giving significant leverage of peers closer to the AP. Conversely, the dynamic topology of FSS reduces such influence.

A case of more direct control was introduced in the ridesourcing case, where an external agent controls the incentive mechanisms in the federation, with the goal of making the federation sustainably grow while gathering benefits. In this case, failure scenarios were avoided by using adequate pricing control actions.

9.2.4 Synergy saturation and federation limits

In all of the cases analyzed here, the distributed synergy of the federations saturates at some point after emergence. This is a universal result originated by technological and physics limits, and bounded stakeholder needs.

For FSS and also for WCN, saturation stems from the finite bandwidth available, and the physical impossibility to have latency performances better than 0. The effect of the network topology, and network edge capacity, also restraints the achievable benefits of cooperation. For ridesourcing, saturation stems from the market parameters and the irreducible vehicle operating costs.

Based on these observations, we here identify 3 types of bounds than constrain the achievable synergy. The first and upper bound to all benefits is the *tradespace range*. The second is the *stakeholder utility range*, and third, the performance bounds originated by *technical limits*.

The most general bound that establishes the absolute maximum achievable NBC by a system is the *tradespace range*. The difference in cost between the non-dominated Pareto architectures that give the maximum and minimum utility is the maximum achievable NBC by any particular system, and hence, their maximum contribution to the overall synergy. That is, the cost range constrains the maximum advantages achievable by federation. Figure 83 illustrates this notion graphically.

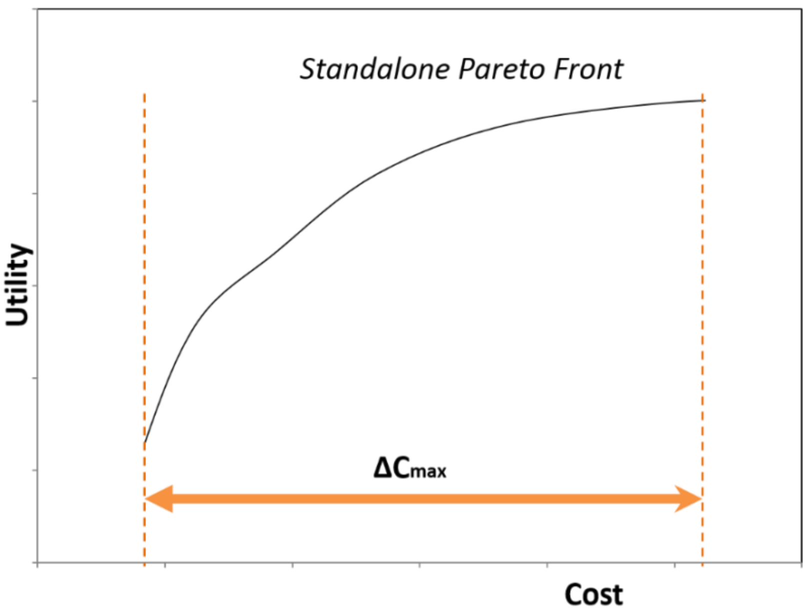

Figure 83. The difference between the maximum utility and the minimum utility architecture establishes the upper bound for the system's NBC and hence FoS synergy.

In the case studies here, we have generated tradespaces covering the largest utility spans possible to have a complete characterization of the design space from utility 0 to 1. When it is possible to explore architectures very close to cost 0, and architectures very close to utility 1, then the *tradespace* range does not restrain significantly the achievable synergy.

In the tradespace we assigned utility 1 to the highest performing architectures. Hence, the *stakeholder utility range* used also bounds the achievable benefits of cooperation. The mapping of performance to stakeholder perceived utility in the [0,1] range establishes a cap for the benefits of performance increases. This is a feature required to realistically represent finite stakeholder needs.

Finally, limits to performance emerge from the nature of the cooperation processes and the physics involved. This can, and does, restrain synergy further. As Figure 84 illustrates, in some situations the maximum utility bound (utility 1) cannot be reached due a performance bound. As mentioned above, performance bounds stem from FoS network topology and the network edges capacity. Moreover, the cooperation might be limited by what we called sharing constraint in chapter 6, that is, the extent of cooperation each system is ready to engage with, in terms of resources offered.

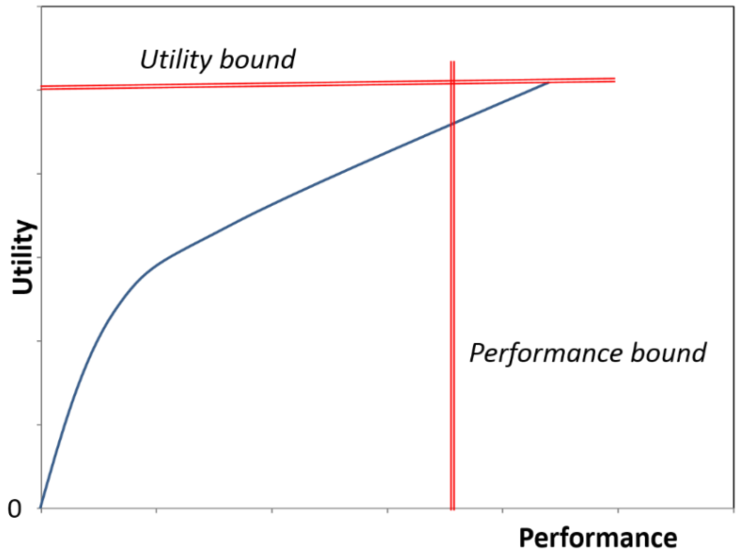

Figure 84. Bounds on utility, that exists in the range [0,1] and performance-induced bounds.

In the case of ridesourcing, which does not need of using the SPF to capture the benefits of cooperating, we can directly use Eq. 7-1 to directly infer the limits of synergy without considering the bounds mentioned above. For instance, that the maximum rider NBC is 14.3 USD. This is the cost advantage of having an average-length, free ride with 0 minutes waiting time, with the standard NYC taxicab parameters.

9.3 Limitations

In chapter 5, we discussed the theoretical limitations of this approach. This included the *non-simultaneity of architecting efforts, the usage of the Pareto front as bijective utility-cost function, the indifference to architectural alternatives, and the rationality of system architects.* These limitations were deemed acceptable for the analysis intended.

Moreover, in the development of the studies herewith we assumed we can characterize the tradespace and utilities of *future* systems and peers.

In the case of WCN we need to assume the geographic, market and technology conditions do not vary much during the network adoption timeframe. This is acceptable for a deployment time of months up to a few years, a typical timeframe for WCNs. Nevertheless, should we have a specific indication of how should these conditions change, it would be straightforward to model peer adoption from a specific time point with the new conditions.

In the case of ridesourcing, a dramatic change in market conditions –as posed by radical changes in urban transport systems, costs of living or transportation– could compromise the predictions of the framework. However such changes are an exception rather than a rule.

For FSS, the mission commissioning cadence implied in the analysis –1 year– does make the framework extend assumptions on mission needs and tradespaces up to 10 years ahead in time. However, given the long development time of space missions, it is possible to know in advance which missions are in the pipeline and what is their general design space and objectives. Publicly available databases and agency resources include details of missions planned up to 15 years in advance (World Meteorological organization, 2015). Moreover, given the current relative rigidness of spacecraft design, their mission goals and stakeholder needs do not change significantly during mission lifetime.

Hence we can conclude the assumptions on future utility functions and tradespaces are reasonable for the cases here and are readily applicable to most fields, excepting the ones undergoing drastic technological changes. Therefore it is not recommended to apply the techniques here to very long timeframes –15 years or more– due the potential for injection of radically new technologies in the systems, operational concepts, or even potential disappearance of the class of systems analyzed.

Shall we want to address the evolution of a federation under significant technological or stakeholder objectives change during the federation deployment, we only need to change the assumptions for utility and tradespaces as the MDP advances. This is naturally leads to the combination of the present framework with other techniques in the literature, like TDN and epoch-era analysis, that have been thoroughly introduced in chapter 4.

Either dynamic or static, the knowledge of the stakeholder performance-utility function details is the most important assumption in the framework. Hence, we have studied extensively the effects of changing utility functions on the results. Changes in shape of the utility function, such as switching from concave to convex, using step functions, and slight variations in values, as reported in the FSS and WCN case, have a relatively low impact (5-20% of synergy) in the appeal and conditions for federation adoption. Instead, when changing the functions regime and adding large plateaus, as reported in in figures 45-48, we were able to exert a significant change in federation RoI, of up to 50%.

The insights from this are that the exact shape of the utility-performance response is not a game-changer if it is positive and monotonic, that is, if there is an effective appetite for additional performance and changes in the latter can be readily captured by the stakeholder. If the stakeholder is insensitive, completely or partially, to performance increases, naturally the benefits rendered by federation are not that appealing.

Finally, the results' precision is also naturally bounded by the underlying systems performances model, which is different for every system. For each case study we have included the specific modelling assumptions and their effect on results, if any. When possible to forecast so, the modelling assumptions in this work attempt to be conservative towards federation benefits.

After understanding their limitations, we can now extract from the results of this work a series of preliminary recommendations for the establishment of federations.

9.4 Recommendations for FoS implementation and governance

The technological and context conditions studied in this thesis for federation can inform the agents faced with the challenges of deploying a federation, being individuals, organizations, and/or system architects. We now distinguish a series of practices to consider.

Generate mechanisms for positive returns to the pioneering peers. The first satellites in a FSS, or the first riders in a ridesourcing market, need to be able to capture the positive returns of their adoption of the federation, either immediately or with delayed but plausible mechanisms. Early ridesourcing peers are not able to capture, at individual level, the benefits of helping to establish a ridesourcing community, neither can sustain the necessary losses for months or years. Therefore these types of peers need to be incentivized immediately. In the case of FSS, due to the significant upfront investment made in the system manufacturing, their long lifetimes, and the limited amount of systems present –global EO just counts a few hundreds of satellites– the systems are capable of capturing future benefits. Therefore they can trade the initial losses for potential rewards downstream, shall those be credible.

To make those rewards credible, besides the application of the framework here that justifies federation on economic grounds, the systems stakeholder would also require certain assurance from other systems, or at least evidence of their intentions, to minimize the risks, when adoption implies significant expenses. Hence, a preliminary federation agreement and a guiding charter would greatly foster the future of the FoS. Another option is to artificially create a FoS within a portfolio of systems owned by the same stakeholder, as we reason next.

Create an FoS internal to an organization and open it. For satellite systems, this means federating several missions within the portfolio of an agency or national government. For WCN, it involves a neighbour community. For resourcing it is more difficult to envision, but we could assimilate it to starting a ridesharing service within a large company and then expanding. This helps overcome the initial negative synergy phases and indeed allows to 'design' the federation nearly as a directed SoS. When the FoS is operational, it can be opened to additional parties to join, hence increasing its values and potentially earning commercial returns for the first systems.

Make peer operations and capabilities available to other peers. This recommendation concerns one of the basic elements of FoS governance discussed in section 5. This basic level of coordination seems absolutely essential. While during day-to-day operation it is acceptable to have dynamics on resources and capabilities shared, realistic expectations are needed at design stage for any peer to realistically consider federation. The urgency of this need escalates with the relative value of the expenses involved and the switching costs. Shall a FoS fail, a ridesourcing rider can simple delete the platform app from their phone, while a driver would need to reconsider their occupation, and a satellite is stuck with a useless payload for all its lifetime. While other levels of governance and control, such as boundary control and incentives management, might be beneficial but not necessary required, a realistic expectation on what to expect from the other peers is fundamental.

We argued in chapter 8 that FoS might be capable of self-regulation in *tragedy of the commons* cases, as the abandoning of the FoS by some systems might return it to a favourable configuration. However, a certain control of the cooperation protocols and resource allocation might be required to keep the FoS in a functional state, and well above margins, for all participants. Additionally, self-regulation will not work when malicious nodes and intentional misbehavior is present. This has been explored in other work (Korobova et al., 2015) and it applies as our discussion here as an argument for adding boundary controls to the FoS governance mechanisms, or for including the detection and suppression of malicious peers in the cooperation protocol.

9.5 Conclusions summary

Figure 85 outlines the various findings listed in this chapter, and the recommendations we have proposed for FoS governance and implementation. We also highlight here the connections between governance strategies, FoS emergence and collapse, types of synergy achievable and its drivers and constraints. This overview and conceptual connection of the different areas that affect the establishment and evolution of a federation is a first attempt at establishing a general theory on FoS.

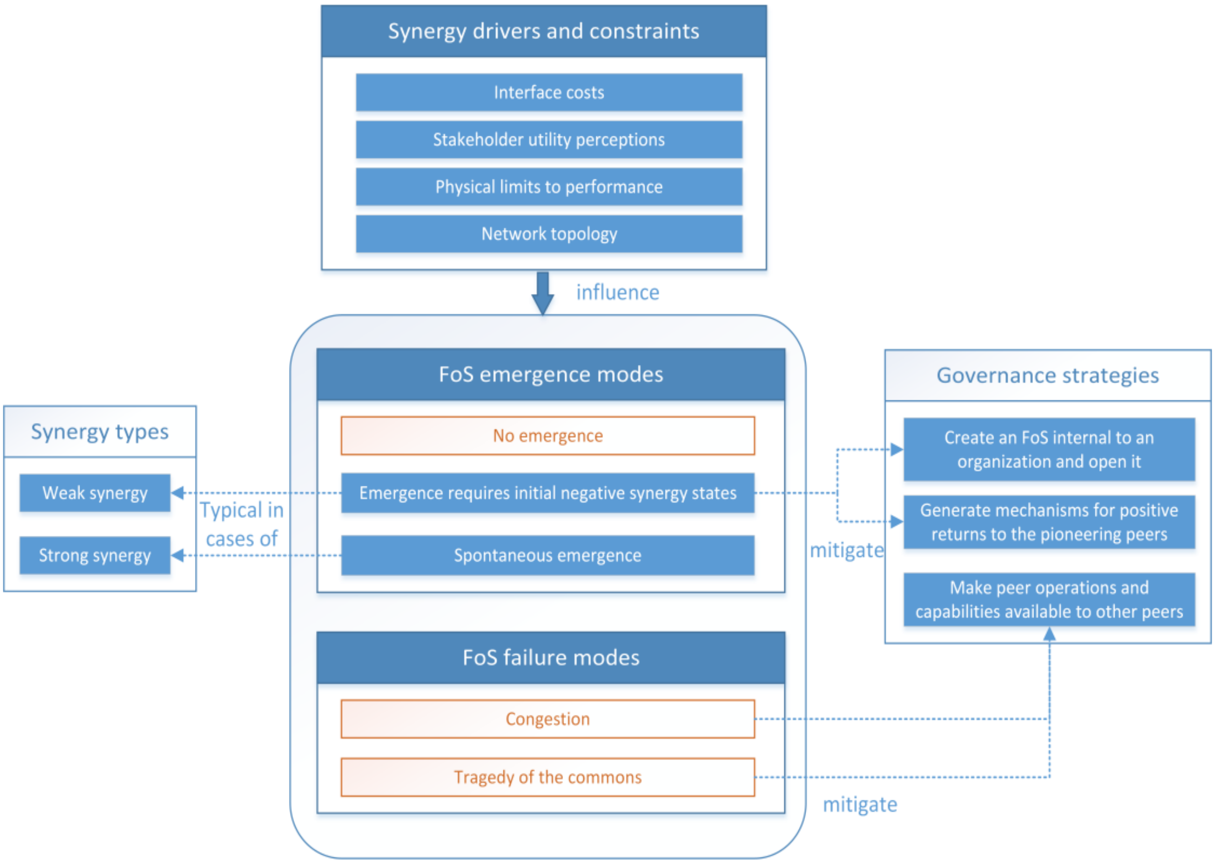

Figure 85. Summary of theoretical findings and recommendations, and their connections.

9.6 Contributions summary and relevance

As mentioned in Chapter 4, the architecting techniques in the literature are not completely adequate, in terms of focus, perspective and model, to analyze FoS. Moreover, FoS are a class of SoS, –virtual SoS– which lack dedicated research efforts so far. While directed SoS have been already the subject of academic and industrial research, collaborative and virtual SoS have not been addressed at large by the systems architecting and systems engineering community. The reasons stem from the inherent lack of global decision makers, and the novelty of such systems, which in some cases depend on relatively new communication technologies.

Therefore, this thesis' contributions are multi-faceted, covering theoretical and practical aspects. First of all, this thesis defines FoS in detail and identifies their existential challenges. In the author's perspective, the most relevant theoretical contribution of this work is the definition and consolidation of FoS as type of SoS.

Second, we have showcased a quantitative analysis in the field of SoS, which in general is short of such examples. More specifically, this thesis contributes to the systems architecting discipline with a new integrated framework to address FoS. Third, we have introduced in the systems engineering and architecting field the concept of synergy and connected it to the concept of emergence and resource allocation. Finally, on the application side, we have provided insights in specific case studies, including satellite systems, transportation and communications.

The approach used in this thesis integrates several methods drawn from systems architecting, operations research and computer science in a cohesive framework, as outlined in Chapter 5. Hence, on methodological grounds, the novelty does not reside on the development of a specific mathematical method but on the tailoring of several to build a novel guideline for engineering practitioners.

Nevertheless, this thesis' faced the issues of the combinatorial explosion of design options that plagues tradespace exploration and many other preliminary design efforts. This issue is even more acute for SoS, which naturally include more decisions to explore. The conventional method to address this issue is to reduce the amount of options to analyze, for instance using DoE. We have here instead exploited the temporal structure of the problem to assess only the promising FoS evolution paths. Moreover, in the FSS study, we have illustrated the finding of a global optimum within a large combinatorial problem, via the formulation of a search problem in a space of FoS states.

Besides the tangible contributions of this thesis, the work here also opens the potential for several other topics to be investigated.

9.7 Future lines of work

Building upon this first piece of work in FoS, we can foresee several future research lines. The technological enablers of FoS, especially regarding communications and in particular MANETs, and SDRs (see section 4.5.1) will continue to evolve on its own and provide affordable and performing interfaces for FoS. One of the most interesting aspects of SDRs for FoS is the possibility to add backwards compatibility in the system, that is, be able to add to the federation systems that already are in operation without modifying their design, but only their operations. This motivated the idea of 'negotiator' nodes in FSS, systems that perform protocol translation and act as middle-men to enable cooperation between two other systems a priori incompatible. Such work is in progress (Akhtyamov et al., 2015). Also concerning FSS, ad-hoc, multi-stakeholder space networks are not yet a mature technology. In contrast, technical enablers for ridesourcing and WCNs are already available and drove the advent of these FoS. Therefore, the work on this thesis aims to pave the way and stimulate the ongoing work in space networks technology.

Another technological area than affects the architecting and specially the mutual trust between peers, both on ground and in space, is the security of the transactions and the handling of malicious nodes. Initial work on this topic has been performed by Korobova (2015).

The establishment of peer trust and federate governance are important components of the success of a federation. Besides the technical and economic conditions for federation emergence identified in this work, the systems' stakeholders, architects and operators also need to trust their federates to provide fair and predictable exchanges of resources, under rules known and accepted by all parties. Mechanisms from the sharing economy and also the lessons for WCNs organization can serve as a model for engineering systems. There is opportunity for additional work on this aspect. Nevertheless, advanced cooperation protocols with embedded provision of security and trust could make peer-to-peer, decentralized governance possible.

Finally, this work scope covers the strategic and lifecycle aspects of federating, but not the tactical, short term organization of exchanges in the federation. We used, when needed, optimization techniques to be able to compare the FoS configurations in a consistent manner, and we captured the basics of offer and demand in the ridesourcing models, but we did not delve into the details of the competition and bid for resources within the federation. Initial work on this aspect has been performed by other authors (Pica and Golkar, 2015). In connection to the short-term organization of the federation, the work of Matevosyan (2015) built a scheduling system on top of FSS to generate ad-hoc, on demand virtual satellite missions (VSM) based upon instrument time of the federates.

In terms of implementation, FoS are in their infancy. As we defined them here, they were hardly possible before the advent of embedded computing power, automation and communication networks. However, we will see more and more systems interconnected in the future, as the concepts of IoT, Smart Cities, Industry 4.0, autonomous cars and intelligent transport flourish. This will potentially form new FoS. A wealth of research is dedicated to the different core technologies enabling these paradigms, but we will also need from advances in systems architecting and engineering practices to support design and decision-making in such complex environments. This thesis intends to contribute to this future and vision.

Chapter 10 References

Abbas, A.E., 2009. Multiattribute utility copulas. Oper. Res. 57, 1367–1383.

Abdelaal, A., Ali, H., 2007. A typology for community wireless networks business models. AMCIS 2007 Proc. 308.

Abdelaal, A., Ali, H., Khazanchi, D., 2009. The role of social capital in the creation of community wireless networks, in: System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on. IEEE, pp. 1–10.

Adams, K., 2015. Non-functional requirements in systems analysis and design. Springer, [S.l.].

Agatz, N., Erera, A., Savelsbergh, M., Wang, X., 2012. Optimization for Dynamic Ride-Sharing: A Review. Eur. J. Oper. Res. 223, 295–303. doi:10.1016/j.ejor.2012.05.028

Agatz, N.A.H., Erera, A.L., Savelsbergh, M.W.P., Wang, X., 2011. Dynamic Ride-Sharing: A Simulation Study in Metro Atlanta. Transp. Res. Part B Methodol. 45, 1450–1464. doi:10.1016/j.trb.2011.05.017

Akhtyamov, R., Cruz, I.L. i, Matevosyan, H., Knoll, D., Pica, U., Lisi, M., Golkar, A., 2016. An implementation of Software Defined Radios for federated aerospace networks: Informing satellite implementations using an inter-balloon communications experiment. Acta Astronaut. doi:10.1016/j.actaastro.2015.12.043

Akhtyamov, R., Lluch, Ignasi, Golkar, Alessandro, 2015. Software Defined Radio Implementation of a Negotiator Node Testbed for Federated Satellite Systems. Presented at the 66th International Astronautical Congress, Jerusalem, Israel.

Aldhaban, F., 2012. Exploring the adoption of Smartphone technology: Literature review, in: 2012 Proceedings of PICMET '12: Technology Management for Emerging Technologies. pp. 2758–2770.

Aliakbargolkar, A., 2012. A framework for space systems architecting under stakeholder objectives ambiguity. Massachusetts Institute of Technology.

Alley, J.K., 2016. The Impact of Uber Technologies on the New York City Transportation Industry.

Amey, A., Attanucci, J., Mishalani, R., 2011. Real-time ridesharing: opportunities and challenges in using mobile phone technology to improve rideshare services. Transp. Res. Rec. J. Transp. Res. Board 103–110.

Anderson, D.N., 2014. "Not just a taxi"? For-profit ridesharing, driver strategies, and VMT. Transportation 41, 1099–1117. doi:10.1007/s11116-014-9531-8

Anderson, D.P., Cobb, J., Korpela, E., Lebofsky, M., Werthimer, D., 2002. SETI@ home: an experiment in public-resource computing. Commun. ACM 45, 56–61.

Andersson, J., 2000. A survey of multiobjective optimization in engineering design. Dep. Mech. Eng. Linktjping Univ. Swed.

Angeletti, P., Lisi, M., Tognolatti, P., 2014. Software Defined Radio: A key technology for flexibility and reconfigurability in space applications, in: 2014 IEEE Metrology for Aerospace (MetroAeroSpace).

Ansoff, H., 1987. The concept of corporate strategy. Homewood IL Irwin.

Armbrust, M., Stoica, I., Zaharia, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., 2010. A view of cloud computing. Commun. ACM 53, 50. doi:10.1145/1721654.1721672

Arrow, K.J., 1966. Exposition of the theory of choice under uncertainty. Synthese 16, 253–269.

Avonts, J., Braem, B., Blondia, C., 2013. A questionnaire based examination of community networks, in: Wireless and Mobile Computing, Networking and Communications (WiMob), 2013 IEEE 9th International Conference on. IEEE, pp. 8–15.

AWMN, 2016. Athens Wireless metropolitan Network [WWW Document]. URL http://www.awmn.net/content.php (accessed 6.30.16).
Baig, R., Roca, R., Freitag, F., Navarro, L., 2015. guifi.net, a crowdsourced network infrastructure held in common. Comput. Netw. 90, 150–165. doi:10.1016/j.comnet.2015.07.009
Ballard, A.H., 1980. Rosette Constellations of Earth Satellites. Aerosp. Electron. Syst. IEEE Trans. On AES-16, 656–673. doi:10.1109/TAES.1980.308932
Bass, F.M., 1969. A new product growth for model consumer durables. Manag. Sci. 15, 215–227.
Bass, F.M., Krishnan, T.V., Jain, D.C., 1994. Why the Bass Model Fits without Decision Variables. Mark. Sci. 13, 203–223. doi:10.1287/mksc.13.3.203
Bertsimas, D., Sim, M., 2004. The price of robustness. Oper. Res. 52, 35–53.
Bertsimas, D., Tsitsiklis, J.N., 1997. Introduction to linear optimization, Athena Scientific series in optimization and neural computation. Athena Scientific, Belmont, Mass.
Beste, D., 1978. Design of Satellite Constellations for Optimal Continuous Coverage. IEEE Trans. Aerosp. Electron. Syst. AES-14, 466–473. doi:10.1109/TAES.1978.308608
Bloomberg News, 2016. Uber Now Most Used Taxi App in 108 Countries [WWW Document]. URL http://adage.com/article/digital/uber-taxi-app-108-countries/305575/
Boardman, J., Sauser, B., 2006. System of Systems-the meaning of of, in: System of Systems Engineering, 2006 IEEE/SMC International Conference on. IEEE, p. 6–pp.
Botsman, R., 2013. The Sharing Economy Lacks a Shared Definition: giving meaning to the terms. Collab. Lab Online Present. Httpwww Slideshare NetCollabLabshared-Def-Pptf.
Bristow, J., Folta, D., Hartman, K., 2000. A formation flying technology vision, in: AIAA Space 2000 Conference and Exposition.
Brown, O., Eremenko, P., 2006. The Value Proposition for Fractionated Space Architectures, in: Space 2006, SPACE Conferences & Exposition. American Institute of Aeronautics and Astronautics.
Cachon, G.P., Daniels, K.M., Lobel, R., 2016. The role of surge pricing on a service platform with self-scheduling capacity.
Camerer, C., Babcock, L., Loewenstein, G., Thaler, R., 1997. Labor supply of New York City cabdrivers: One day at a time. Q. J. Econ. 112, 407–441.
Car 2 Car communication Consortium, 2011. Memorandum of Understanding for OEMs within the CAR 2 CAR Communication Consortium on Deployment Strategy for cooperative ITS in Europe.
Cascetta, E., 2009. Transportation Systems Analysis, Springer Optimization and Its Applications. Springer US, Boston, MA. doi:10.1007/978-0-387-75857-2
Castells, M., 2011. The rise of the network society: The information age: Economy, society, and culture. John Wiley & Sons.
Chan, N.D., Shaheen, S.A., 2012. Ridesharing in North America: Past, Present, and Future. Transp. Rev. 32, 93–112. doi:10.1080/01441647.2011.621557
Chatterjee, S., 1986. Types of synergy and economic value: The impact of acquisitions on merging and rival firms. Strateg. Manag. J. 7, 119–139.
Chattopadhyay, D., 2009. A method for tradespace exploration of systems of systems. Massachusetts Institute of Technology.
Chattopadhyay, D., Ross, A.M., Rhodes, D.H., 2009. Demonstration of system of systems multi-attribute tradespace exploration on a multi-concept surveillance architecture, in: 7th Conference on Systems Engineering Research.
Chattopadhyay, D., Ross, A.M., Rhodes, D.H., 2008. A framework for tradespace exploration of systems of systems, in: 6th Conference on Systems Engineering Research, Los Angeles, CA.
Chen, M.K., Sheldon, M., 2015. Dynamic pricing in a labor market: Surge pricing and flexible work on the Uber platform. Mimeo, UCLA.
Chieng, D., Ting, A., 2011. Design and analysis of WiMAX-WiFi multi-tier network, in: The International Conference on Information Networking 2011 (ICOIN2011). pp. 388–393. doi:10.1109/ICOIN.2011.5723115
Chinchuluun, A. (Ed.), 2008. Pareto optimality, game theory and equilibria, Springer optimization and its applications. Springer, New York.
Collopy, P., 2009. Aerospace System Value Models: A Survey and Observations. American Institute of Aeronautics and Astronautics. doi:10.2514/6.2009-6560

Copernicus Space Component Data Access (CSCDA), n.d. Copernicus Contributing Missions [WWW Document]. URL https://spacedata.copernicus.eu/web/cscda/missions (accessed 4.20.16).
Crawley, E., Cameron, B., Selva, D., 2016a. System architecture: strategy and product development for complex systems. Pearson, Boston.
Crawley, E., Cameron, B., Selva, D., 2016b. System architecture: strategy and product development for complex systems. Pearson, Boston.
Crawley, E., De Weck, O., Magee, C., Moses, J., Seering, W., Schindall, J., Wallace, D., Whitney, D., others, 2004. The influence of architecture in engineering systems (monograph.
Dahmann, J., Rebovich, G., Lowry, R., Lane, J., Baldwin, K., 2011. An implementers' view of systems engineering for systems of systems, in: Systems Conference (SysCon), 2011 IEEE International. IEEE, pp. 212–217.
Dan Leone, 2012. TDRS-K Launch Caught Up In Cascade of Fla. Delays.
Davendralingam, N., DeLaurentis, D., 2013. A robust optimization framework to architecting system of systems. Procedia Comput. Sci. 16, 255–264.
Dawes, M., 2016. Perspectives on the Ridesourcing Revolution: surveying individual attitudes toward Uber and Lyft to inform urban transportation policymaking. Massachusetts Institute of Technology.
De Weck, O.L., 2004. Multiobjective optimization: History and promise, in: Invited Keynote Paper, GL2-2, The Third China-Japan-Korea Joint Symposium on Optimization of Structural and Mechanical Systems, Kanazawa, Japan. p. 34.
De Weck, O.L., Roos, D., Magee, C.L., 2012. Engineering systems: meeting human needs in a complex technological world. MIT Press, Cambridge, Mass.
de Weck, O.L., Ross, A.M., Rhodes, D.H., 2012. Investigating relationships and semantic sets amongst system lifecycle properties (ilities), in: Third International Engineering Systems Symposium CESUN. Citeseer, pp. 18–20.
del Portillo, I., Bou-Balust, E., Alarcon, E., 2015. Scalability of Resource Constrained General Purpose Fractionated Satellite Network Architectures. Presented at the IEEE Aerospace Conference, Montana, US.
DeLaurentis, D.A., 2005. A taxonomy-based perspective for systems of systems design methods, in: Systems, Man and Cybernetics, 2005 IEEE International Conference on. IEEE, pp. 86–91.
Delosieres, L., Nadjm-Tehrani, S., 2012. BATMAN store-and-forward: The best of the two worlds, in: Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on. pp. 721–727. doi:10.1109/PerComW.2012.6197608
Denhart, J., Gemmer, T., Ferguson, S., Mazzoleni, A., 2013. Tradespace Exploration of Reconfigurability using Surrogate Points to Simplify Dominance Determination. Presented at the 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference.
Dervojeda, K., Verzijl, D., Nagtegaal, F., Lengton, M., Rouwmaat, E., Monfardini, E., Frideres, L., 2013. The sharing economy: accessibility based business models for peer-to-peer markets. Eur. Comm. Bus. Innov. Obs. Sept.
Dimitrakopoulos, G., Demestichas, P., 2010. Intelligent Transportation Systems. IEEE Veh. Technol. Mag. 5, 77–84. doi:10.1109/MVT.2009.935537
DoD, 2008. Systems Engineering Guide for Systems of Systems. US Department of Defense, Director, Systems and Software Engineering, Deputy Under Secretary of Defense (Acquisition and Technology), Office of the Under Secretary of Defense (Acquisition, Technology and Logistics).
Don, C., Sage, A.P., 2012. A framework for interoperability assessments in Systems of Systems and Families of Systems. Inf. Knowl. Syst. Manag. 275–295. doi:10.3233/IKS-2012-0211
Edwards, B.L., 2014. NASA'S CURRENT ACTIVITIES IN FREE SPACE OPTICAL COMMUNICATIONS, in: International Conference on Space Optics. p. 10.
Eilertsen, B., 2012. Ground station networks vs. GEO relay satellite systems for polar orbiting satellites, in: SpaceOps 2012. p. 1290826.
Eisner, H., 2008. Essentials of project and systems engineering management. John Wiley & Sons.
Engelen, S., Gill, E.K., Verhoeven, C.J., 2011. Systems engineering challenges for satellite swarms, in: Aerospace Conference, 2011 IEEE. IEEE, pp. 1–8.

Eoportal missions directory [WWW Document], 2016. URL https://eoportal.org/web/eoportal/satellite-missions
Eurona sat, 2017. Internet Rural Satelite [WWW Document]. URL http://www.internet-satelite.eu/ (accessed 5.15.17).
European Comission, 2016. Broadband Coverage in Europe 2015. EC DG Communications Networks, Content and Technology, Brussels.
European Comission, 2013. A Digital Agenda For Europe.
European Commission, Directorate-General for the Information Society and Media, SamKnows Limited, 2015. Quality of broadband services in the EU: final report : October 2014. Publications Office, Luxembourg.
Evangelista, R., Guerrieri, P., Meliciani, V., 2014. The economic impact of digital technologies in Europe. Econ. Innov. New Technol. 23, 802–824.
Faber, N., Nakamura, Y., Alena, R., Mauro, D., Frost, C.R., Bhat, G., McNair, J., 2014. Heterogeneous Spacecraft Networks: General concept and case study of a cost-effective, multi-institutional Earth observation platform, in: Aerospace Conference, 2014 IEEE. pp. 1–16. doi:10.1109/AERO.2014.6836297
Farber, H.S., 2005. Is tomorrow another day? The labor supply of New York City cabdrivers. J. Polit. Econ. 113, 46–82.
Federal Communications Provision, 2017. Broadband speed guide [WWW Document]. URL https://www.fcc.gov/reports-research/guides/broadband-speed-guide (accessed 5.16.17).
Feinberg, E.A., Shwartz, A., 2012. Handbook of Markov decision processes: methods and applications. Springer Science & Business Media.
Festag, A., 2014. Cooperative intelligent transport systems standards in europe. IEEE Commun. Mag. 52, 166–172. doi:10.1109/MCOM.2014.6979970
Fleisch, E., 2010. What is the Internet of Things. Things Add Value Auto-ID Labs White Pap. WP-BIZAPP-053 Auto-ID Lab St Gallen Switz. 1.
Fleisher, L., 2014. Thousands of European Cab Drivers Protest Uber, Taxi Apps: Protesters in London, Madrid, Milan Say the Apps Skirt Regulations. Wall Str. J.
Flores-Guri, D., 2003. An economic analysis of regulated taxicab markets. Rev. Ind. Organ. 23, 255–266.
Ford Jr, L., Fulkerson, D.R., 1956. Solving the transportation problem. Manag. Sci. 3, 24–32.
Frank J. Stocklin, Michal W. Rackley, Ted C. Sobchak, 2012. Space Network Users' Guide (SNUG) (No. revision 10). Goddard Space Flight Center, Greenbelt, Maryland.
Fricke, E., Schulz, A.P., 2005. Design for changeability (DfC): Principles to enable changes in systems throughout their entire lifecycle. Syst. Eng. 8, no–no. doi:10.1002/sys.20039
Fundació Privada per a la Xarxa Oberta, Lliure i Neutral Guifi.net, 2017. Guifi.Net webpage [WWW Document]. URL https://guifi.net/ (accessed 6.30.16).
Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.-E., Wang, X., Koenig, S., 2013. Ridesharing: The state-of-the-art and future directions. Transp. Res. Part B Methodol. 57, 28–46. doi:10.1016/j.trb.2013.08.012
Gastner, M.T., Newman, M.E., 2006. The spatial structure of networks. Eur. Phys. J. B-Condens. Matter Complex Syst. 49, 247–252.
Gerla, M., Lee, E.-K., Pau, G., Lee, U., 2014. Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds. IEEE, pp. 241–246. doi:10.1109/WF-IoT.2014.6803166
Grau, J.M.S., Romeu, M.E., 2012. A mathematical approach to the modeling of taxi services, in: European Symposium on Quantitative Methods in Transportation Systems. "LATSIS 2012 (1st European Symposium on Quantitative Methods in Transportation Systems)." Lausane.
Griffith, V., Koch, C., 2014. Guided Self-Organization: Inception, in: Prokopenko, M. (Ed.), . Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 159–190.
Grogan, P.T., Golkar, A., Shirasaka, S., de Weck, O.L., 2014. Multi-stakeholder interactive simulation for federated satellite systems, in: Aerospace Conference, 2014 IEEE. pp. 1–15. doi:10.1109/AERO.2014.6836253
Gunter, B.C., Maessen, D.C., 2013. Space-based distributed computing using a networked constellation of small satellites. J. Spacecr. Rockets 50, 1086–1095.
Ha, J.C. (Ed.), 1998. Mission Design & Implementation of Satellite Constellations, Space Technology Proceedings. Springer Netherlands, Dordrecht.

Haider, M., 2015. To Uber or Not to Uber: That Is the Question.
Hall, J.V., Krueger, A.B., 2016. An analysis of the labor market for Uber's driver-partners in the United States. National Bureau of Economic Research.
Handy, C., 1992. Balancing corporate power: A new federalist paper. Harv. Bus. Rev.
Harding, S., Kandlikar, M., Gulati, S., 2016. Taxi apps, regulation, and the market for taxi journeys. Transp. Res. Part Policy Pract. 88, 15–25. doi:10.1016/j.tra.2016.03.009
Harris, M.A., Thoman, B.E., Carter, D.L., 2016. Near Earth Network (NEN) User's Guide (No. revision 2). Goddard Space Flight Center, Greenbelt, Maryland.
Hartmans, A., 2016. Uber says it has over 80% of the ride-hailing market in the U.S [WWW Document]. Uber Says It Has 80 Ride-Hailing Mark. US. URL http://www.businessinsider.com/uber-majority-ride-hailing-market-share-lyft-us-2016-8?utm_source=feedly&utm_medium=webfeeds (accessed 5.2.17).
Hemmati, H., 2006. Deep space optical communications. John Wiley & Sons.
Herberg, U., Clausen, T., Milan, J., 2010. Digital Signatures for Admittance Control in the Optimized Link State Routing Protocol Version 2, in: Internet Technology and Applications, 2010 International Conference on. pp. 1–4. doi:10.1109/ITAPP.2010.5566285
Hoebeke, J., Moerman, I., Dhoedt, B., Demeester, P., 2004. An overview of mobile ad hoc networks: Applications and challenges. J.-Commun. Netw. 3, 60–66.
Holler, J., Tsiatsis, V., Mulligan, C., Avesand, S., Karnouskos, S., Boyle, D., 2014. From Machine-to-machine to the Internet of Things: Introduction to a New Age of Intelligence. Academic Press.
Holodny, E., 2016. Uber and Lyft are demolishing New York City taxi drivers. Bus. Insid.
Huang, Y., Kannan, G., Bhatti, S., Merchant, S.N., Desai, U.B., 2008. Route dynamics for Shortest Path First routing in mobile ad hoc networks, in: Wireless Telecommunications Symposium, 2008. WTS 2008. pp. 236–242. doi:10.1109/WTS.2008.4547570
IDESCAT, 2016. Densidad de población. Comarcas y Aran, ámbitos y provincias. [WWW Document]. URL http://www.idescat.cat/pub/?id=aec&n=249&lang=es (accessed 5.17.17).
Isaac, E., 2014. Disruptive Innovation: Risk-Shifting and Precarity in the Age of Uber. Berkeley Roundtable Int. Econ. BRIE Work. Pap. 7.
Ishmael, J., Bury, S., Pezaros, D., Race, N., 2008. Deploying rural community wireless mesh networks. IEEE Internet Comput. 12.
ISO/IEC/IEEE International Standard - Systems and software engineering – System life cycle processes, 2015. . ISOIECIEEE 15288 First Ed. 2015-05-15 1–118. doi:10.1109/IEEESTD.2015.7106435
ITU, 2016. ICT facts and figures 2016. ICT data and statistics division, Generva, Switzerland.
Iversen, M., 1997. Concepts of synergy-Towards a clarification, in: Prieiga per Interneta:< Http://Www. Druid. Dk/Conferences/winter1997/Conf-Papers/Mikael. Pdf.
Jamshidi, M. (Ed.), 2009. Systems of systems engineering: principles and applications. CRC Press/Taylor & Francis Group, Boca Raton, FL.
Jiang, S., He, D., Rao, J., 2005. A prediction-based link availability estimation for routing metrics in MANETs. Netw. IEEEACM Trans. On 13, 1302–1312. doi:10.1109/TNET.2005.860094
Johal, S., Zon, N., 2015. Policymaking for the sharing economy: Beyond Whack-a-Mole.
Jon russell, 2015. China's Top Two Taxi-Hailing Services Confirm That They Will Merge [WWW Document]. URL https://techcrunch.com/2015/02/13/kuaidi-dache-didi-dache-merge/ (accessed 5.2.17).
Kankaku, Y., Suzuki, S., Osawa, Y., 2013. ALOS-2 mission and development status, in: Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International. IEEE, pp. 2396–2399.
Karl, H., Willig, A., 2007. Protocols and architectures for wireless sensor networks, 2007 John Wiley & Sons Ltd. pbk. ed. ed. John Wiley & Sons, Chichester, West Sussex, England ; Hoboken, NJ.
Katz, M.L., Shapiro, C., 1985. Network externalities, competition, and compatibility. Am. Econ. Rev. 75, 424–440.
Kingsbury, R., Caplan, D., Cahoy, K., 2015. Compact optical transmitters for CubeSat free-space optical communications, in: Proc. SPIE. p. 93540S.
Korobova, O., Chiesa, A., Golkar, A., 2015. Data authenticity in Federated Satellite Systems. Presented at the 3rd Fractionated and Federated Workshop, Ithaca (NY).

Kurano, M., Yasuda, M., Nakagami, J., Yoshida, Y., 2003. Markov decision processes with fuzzy rewards. J. Nonlinear Convex Anal. 4, 105–116.
Kuttner, H., 2012. Broadband for Rural America: Economic Impacts and Economic Opportunities. Economic Policy/Briefing Paper. Hudson Inst. NJ1.
Lautenbacher, C.C., 2006. The Global Earth Observation System of Systems: Science Serving Society. Space Policy 22, 8–11. doi:10.1016/j.spacepol.2005.12.004
Liu, C., Chau, K.T., Wu, D., Gao, S., 2013. Opportunities and Challenges of Vehicle-to-Home, Vehicle-to-Vehicle, and Vehicle-to-Grid Technologies. Proc. IEEE 101, 2409–2427. doi:10.1109/JPROC.2013.2271951
Lluch, Golkar, 2015. Adopting the Federated Satellite Systems Paradigm for Earth Observation: an Architecting Framework. Presented at the 66th International Astronautical Congress, Jerusalem, Israel.
Lluch i Cruz, I., Golkar, A., 2014. Resource Balancing Analysis of Federated Satellite Systems, in: AIAA SPACE 2014 Conference and Exposition, SPACE Conferences & Exposition. American Institute of Aeronautics and Astronautics. doi:10.2514/6.2014-4270
Lluch, I., Golkar, A., 2015. Design Implications for Missions Participating in Federated Satellite Systems. J. Spacecr. Rockets 1–14. doi:10.2514/1.A33172
Lluch, I., Grogan, P.T., Pica, U., Golkar, A., 2015. Simulating a proactive ad-hoc network protocol for Federated Satellite Systems. IEEE, pp. 1–16. doi:10.1109/AERO.2015.7118984
Luders, R.D., 1961. Satellite Networks for Continuous Zonal Coverage. ARS J. 31, 179–184. doi:10.2514/8.5422
M. Abusali, P.A., Tapley, B.D., Schutz, B.E., 1998. Autonomous Navigation of Global Positioning System Satellites Using Cross-Link Measurements. J. Guid. Control Dyn. 21, 321–327. doi:10.2514/2.4238
Maheshwarappa, M.R., Bowyer, M., Bridges, C.P., 2015. Software Defined Radio (SDR) architecture to support multi-satellite communications, in: 2015 IEEE Aerospace Conference. IEEE, pp. 1–10.
Maier, M.W., 1998. Architecting principles for systems-of-systems. Syst. Eng. 1, 267–284. doi:10.1002/(SICI)1520-6858(1998)1:4<267::AID-SYS3>3.0.CO;2-D
Maier, M.W., 1996. Architecting principles for systems-of-systems, in: INCOSE International Symposium. Wiley Online Library, pp. 565–573.
Maier, M.W., Rechtin, E., 2009. The art of systems architecting, 3rd ed. ed. CRC Press, Boca Raton.
Maisonobe, L., Pommier-Maurussane, V., 2010. Orekit: an Open-source Library for Operational Flight Dynamics Applications. 4th ICATT.
Malhotra, A., Van Alstyne, M., 2014. The dark side of the sharing economy… and how to lighten it. Commun. ACM 57, 24–27.
Maral, G., 2009. Satellite communications systems: systems, techniques and technology, 5th ed. ed. John Wiley, Chichester, West Sussex, U.K.
Markides, C.C., Williamson, P.J., 1994. Related diversification, core competences and corporate performance. Strateg. Manag. J. 15, 149–165.
Marler, R.T., Arora, J.S., 2004. Survey of multi-objective optimization methods for engineering. Struct. Multidiscip. Optim. 26, 369–395. doi:10.1007/s00158-003-0368-6
Martel, J., 1981. NAVSTAR GLOBAL POSITIONING SYSTEM, in: Space Tracking and Data Systems; Proceedings of the Symposium, Meeting Paper Archive. American Institute of Aeronautics and Astronautics.
Martin, P.K., 2014. Space Communications and Navigation: NASA's Management of the Space Network.
Matevosyan, H., Taylor, C.E., Golkar, A., 2015. Evaluating Virtual Satellite Mission Opportunities, in: AIAA SPACE 2015 Conference and Exposition. p. 4674.
Mausam, Kolobov, A., 2012. Planning with Markov Decision Processes: An AI Perspective. Synth. Lect. Artif. Intell. Mach. Learn. 6, 1–210. doi:10.2200/S00426ED1V01Y201206AIM017
McFadden, D., 1974. The measurement of urban travel demand. J. Public Econ. 3, 303–328.
Mekdeci, B., Shah, N., Ross, A.M., Rhodes, D.H., Hastings, D., 2014. Revisiting the Question: Are Systems of Systems just (traditional) Systems or are they a new class of Systems? Syst. Eng. Adv. Res. Initiat. SEAri Camb. MA.
Meyers, L., 2015. Towards a Cost Estimate of a NYC UberX Driver.
Mitola, J., 1992. Software radios-survey, critical evaluation and future directions. IEEE, p. 13/15-13/23. doi:10.1109/NTC.1992.267870

Mollick, E., 2014. The dynamics of crowdfunding: An exploratory study. J. Bus. Ventur. 29, 1–16. doi:10.1016/j.jbusvent.2013.06.005
Monahan, G.E., 1982. State of the art—a survey of partially observable Markov decision processes: theory, models, and algorithms. Manag. Sci. 28, 1–16.
Muri, P., McNair, J., 2012. A survey of communication sub-systems for intersatellite linked systems and CubeSat missions. JCM 7, 290–308.
Narzisi, G., 2008. Multi-Objective Optimization. Quick Introd. N. Y. Univ. Lect.
NASA, 2015a. FY16 NEN reimbursable rates.
NASA, 2015b. FY16 Space Network Reimbursable rates.
NASA, A., 2007. Nasa systems engineering handbook. Wash. DC NASA Cent. Aerosp. Inf.
NASA kennedy, 2015. NASA Awards Launch Services Contract for Next Tracking, Data Relay Satellite [WWW Document]. URL https://www.nasa.gov/press-release/nasa-awards-launch-services-contract-for-next-tracking-data-relay-satellite (accessed 4.10.17).
New York City Taxi &Limousine Comission, 2016. 2016 Taxicab factbook. New York, NY.
New York City Taxi &Limousine Comission, 2014. 2014 Taxicab factbook. New York, NY.
Newcomer, E., 2016. Uber Loses at Least $1.2 Billion in First Half of 2016 [WWW Document]. URL https://www.bloomberg.com/news/articles/2016-08-25/uber-loses-at-least-1-2-billion-in-first-half-of-2016 (accessed 2.5.16).
Nilchiani, R., 2005. Measuring space systems flexibility: a comprehensive six-element framework. Massachusetts Institute of Technology.
Office of the Mayor, City of New York, 2016. For-Hire Vehicle Transportation Study. New York City Government.
Oliver, M., Zuidweg, J., Batikas, M., 2010. Wireless Commons against the digital divide, in: 2010 IEEE International Symposium on Technology and Society. pp. 457–465. doi:10.1109/ISTAS.2010.5514608
ONION consortium, 2016. Operational Network of Individual Observation Nodes [WWW Document]. URL http://cordis.europa.eu/project/rcn/199272_en.html
Ovando, C., Pérez, J., Moral, A., 2015. LTE techno-economic assessment: The case of rural areas in Spain. Telecommun. Policy 39, 269–283. doi:10.1016/j.telpol.2014.11.004
Ovando Chico, M.C., 2015. Aplicación del análisis tecno-económico al despliegue de LTE en zonas rurales de España en la banda de 800MHz (phdthesis). Telecomunicacion.
Palermo, G., Golkar, A., Gaudenzi, P., 2015. Earth Orbiting Support Systems for commercial low Earth orbit data relay: Assessing architectures through tradespace exploration. Acta Astronaut. 111, 48–60.
Pareto, V., 1906. Manuale di economia politica. Societa Editrice.
Penrose, E.T., 1995. The Theory of the Growth of the Firm. Oxford University Press, USA.
Pica, U., Golkar, A., 2015. Sealed-Bid Reverse Auction Pricing Mechanisms for Sharing Economy Systems. INCOSE J. Syst. Eng.
Poghosyan, A., Lluch, I., Matevosyan, H., Lamb, A., Moreno, C., Taylor, C., Golkar, A., Cote, J., Mathieu, S., Pierotti, S., Grave, J., Narkiewicz, J., Topczewski, S., Sochacki, M., Lancheros, E., Park, H., 2016. Unified Classification for Distributed Satellite Systems. Presented at the 4th International Federated and Fractionated Satellite Systems Workshop, Rome, Italy.
Renner, U., Nauck, J., 1984. Development trends in Europe on satellite clusters and geostationary platforms, in: 10th Communication Satellite Systems Conference. pp. 622–628.
Ricci, N., Fitzgerald, M.E., Ross, A.M., Rhodes, D.H., 2014. Architecting systems of systems with ilities: an overview of the SAI Method. Procedia Comput. Sci. 28, 322–331.
Riding, J.L., Ellershaw, J.C., Tran, A.V., Guan, L.-J., Smith, T., 2009. Economics of broadband access technologies for rural areas, in: Optical Fiber Communication-Incudes Post Deadline Papers, 2009. OFC 2009. Conference on. IEEE, pp. 1–3.
Rogers, B., 2015. The Social Costs of Uber. SSRN Electron. J. doi:10.2139/ssrn.2608017
Rogers, E.M., 2001. The digital divide. Convergence 7, 96–111.
Rosello, J., Martellucci, A., Acosta, R., Nessel, J., Braten, L.E., Riva, C., 2012. 26-GHz data downlink for LEO satellites, in: Antennas and Propagation (EUCAP), 2012 6th European Conference on. pp. 111–115. doi:10.1109/EuCAP.2012.6206717
Ross, A., Beesemyer, J., Rhodes, D., 2011. A prescriptive semantic basis for system lifecycle properties. SEAri Working Paper Series, WP-2011-2-1.

Ross, A.M., 2006. Managing unarticulated value: changeability in multi-attribute tradespace exploration. Eng. Syst. Div. 361.
Ross, A.M., Hastings, D.E., 2005. The tradespace exploration paradigm, in: INCOSE International Symposium. Wiley Online Library, pp. 1706–1718.
Ross, A.M., McManus, H.L., Long, A., Richards, M.G., Rhodes, D.H., Hastings, D.E., 2008a. Responsive systems comparison method: Case study in assessing future designs in the presence of change, in: AIAA Space. p. 9.
Ross, A.M., O'neill, M.G., Hastings, D.E., Rhodes, D.H., 2010. Aligning perspectives and methods for value-driven design, in: AIAA Space.
Ross, A.M., Rhodes, D.H., 2008. 11.1. 1 Using Natural Value-Centric Time Scales for Conceptualizing System Timelines through Epoch-Era Analysis, in: INCOSE International Symposium. Wiley Online Library, pp. 1186–1201.
Ross, A.M., Rhodes, D.H., Hastings, D.E., 2008b. Defining changeability: Reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value. Syst. Eng. 11, 246–262.
Russell, S.J., Norvig, P., 2002. Artificial intelligence: a modern approach (International Edition).
Sage, A.P., Cuppan, C.D., 2001. On the Systems Engineering and Management of Systems of Systems and Federations of Systems. Inf Knowl Syst Manag 2, 325–345.
Saleh, J.H., Lamassoure, E., Hastings, D.E., 2002. Space Systems Flexibility Provided by On-Orbit Servicing: Part 1. J. Spacecr. Rockets 39, 551–560. doi:10.2514/2.3844
Schneider, T.W., 2016. Taxi, Uber, and Lyft Usage in New York City [WWW Document]. URL (accessed 2.5.17).
Schulte, F., Voß, S., 2015. Decision Support for Environmental-friendly Vehicle Relocations in Free- Floating Car Sharing Systems: The Case of Car2go. Procedia CIRP 30, 275–280. doi:10.1016/j.procir.2015.02.090
Selva, D., Golkar, A., Korobova, O., Cruz, I.L. i, Collopy, P., de Weck, O.L., 2017. Distributed Earth Satellite Systems: What Is Needed to Move Forward? J. Aerosp. Inf. Syst. 14, 412–438. doi:10.2514/1.I010497
Sender, H., 2014. Here's How Long It Takes to Get an Uber in U.S. Cities. NewsWeek.
Shaheen, S., Stocker, A., Mundler, M., 2017. Online and App-Based Carpooling in France: Analyzing Users and Practices—A Study of BlaBlaCar, in: Disrupting Mobility. Springer, pp. 181–196.
Shaw, G.B., Miller, D.W., Hastings, D.E., 1999. The generalized information network analysis methodology for distributed satellite systems. Citeseer.
Sheldon, M., 2015. Income targeting and the ride-sharing market. Work Pap Univ Chic.
Siddiqi, A., de Weck, O.L., 2008. Modeling methods and conceptual design principles for reconfigurable systems. J. Mech. Des. 130, 101102.
Silver, D., 2015. Markov Decision Processes, in: Lecture Notes on Reinforcement Learning. University College London.
Silver, M.R., De Weck, O.L., 2007. Time-expanded decision networks: A framework for designing evolvable complex systems. Syst. Eng. 10, 167–188.
Slee, T., 2016. What's Yours is Mine: Against the Sharing Economy. Or Books.
Sobieszczanski-Sobieski, J., 2008. Integrated System-of-Systems Synthesis. AIAA J. 46, 1072–1080. doi:10.2514/1.27953
Spaceflight Networks, 2017. SpaceFlight Networks Data Sheet. Seattle, Washington.
Stepp, E., 2014. "Your Driving Costs" study [WWW Document]. URL http://newsroom.aaa.com/2014/05/owning-and-operating-your-vehicle-just-got-a-little-cheaper-aaas-2014-your-driving-costs-study-archive/ (accessed 4.5.17).
Stiglitz, J.E., 2009. The Current Economic Crisis and Lessons for Economic Theory. East. Econ. J. 35, 281–296.
Stiles, J., Goodman, N., Lin, S., 2000. Performance and processing of SAR satellite clusters, in: Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International. IEEE, pp. 883–885.
Sundararajan, A., 2016. The sharing economy: the end of employment and the rise of crowd-based capitalism. The MIT Press, Cambridge, Massachusetts.
Surrey Satellite Technology US, 2017. Surrey X-Band Payload Data Downlink Suite.
Takuto Ishimatsu, Olivier De Weck, Jeffrey Hoffman, Yoshiaki Ohkami, Robert Shishko, 2013. A Generalized Multi-Commodity Network Flow Model for Space Exploration Logistics, in:

AIAA SPACE 2013 Conference and Exposition, SPACE Conferences & Exposition. American Institute of Aeronautics and Astronautics.
Townsend, L., Sathiaseelan, A., Fairhurst, G., Wallace, C., 2013. Enhanced broadband access as a solution to the social and economic problems of the rural digital divide. Local Econ. 28, 580–595. doi:10.1177/0269094213496974
Uber, 2017. Our trip history [WWW Document]. Our Trip Hist. URL https://www.uber.com/en-RU/our-story/ (accessed 2.5.17).
Uber, 2014. Three Septembers of uberX in New York City [WWW Document]. URL https://newsroom.uber.com/us-new-york/three-septembers-of-uberx-in-new-york-city/ (accessed 2.5.17).
Uber Estimate, 2017. NYC Uber Prices, CURRENT & HISTORICAL RATES [WWW Document]. URL http://uberestimate.com/prices/New-York-City/ (accessed 4.5.17).
Ubiquiti Networks, 2017. Ubiquiti Networks Products [WWW Document]. URL https://www.ubnt.com/products/ (accessed 5.17.17).
U.S. Bureau of Labor Statistics, 2016. Occupational Employment and Wages, May 2016 -53-3041 Taxi Drivers and Chauffeurs. United States Department of Labor.
Uy, M., Telford, J.K., 2009. Optimization by Design of Experiment techniques, in: Aerospace Conference, 2009 IEEE. IEEE, pp. 1–10.
van den Steenhoven, Burale, Idil, Toye, Vanssa, Rinne, April, 2016. Shifting Perspectives. Redesigning Regulation for the Sharing Economy. MaRS Solutions lab.
Vanderbei, R.J., 2013. Linear programming: foundations and extensions. Springer, New York.
Vazirani, U., Papadimitriou, C., Dasgupta, S., 2006. Algorithms. McGraw-Hill Education.
Vega, D., Cerdà-Alabern, L., Navarro, L., Meseguer, R., 2012. Topology patterns of a community network: Guifi.net, in: 2012 IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). pp. 612–619. doi:10.1109/WiMOB.2012.6379139
Venkataraman, P., 2009. Applied optimization with MATLAB programming. John Wiley & Sons.
Vergara Pardillo, A., 2011. Aplicación del análisis tecno-económico al despliegue de redes de acceso de próxima generación. El caso de la competencia entre plataformas, la regulación y las políticas públicas en España (phdthesis). Telecomunicacion.
Viscito, L., 2009. Quantifying flexibility in the operationally responsive space paradigm. Massachusetts Institute of Technology.
Vladimirova, T., Wu, X., Bridges, C.P., 2008. Development of a Satellite Sensor Network for Future Space Missions, in: Aerospace Conference, 2008 IEEE. pp. 1–10. doi:10.1109/AERO.2008.4526248
Von Neumann, J., Morgenstern, O., 2007. Theory of games and economic behavior, 60th anniversary ed. ed, Princeton classic editions. Princeton University Press, Princeton, N.J. ; Woodstock.
Walden, D.D., Roedler, G.J., Forsberg, K., Hamelin, R.D., Shortell, T.M., International Council on Systems Engineering (Eds.), 2015. Systems engineering handbook: a guide for system life cycle processes and activities, 4th edition. ed. John Wiley & Sons Inc, Hoboken, New Jersey.
Walker, J.G., 1977. Continuous whole-earth coverage by circular-orbit satellite patterns. DTIC Document.
Wertz, J.R., Everett, D.F., Puschell, J.J. (Eds.), 2011. Space mission engineering: the new SMAD, Space technology library. Microcosm Press : Sold and distributed worldwide by Microcosm Astronautics Books, Hawthorne, CA.
Wertz, J.R., Larson, W.J. (Eds.), 1999. Space mission analysis and design, 3rd ed. ed, Space technology library. Microcosm ; Kluwer, El Segundo, Calif. : Dordrecht ; Boston.
Wikipedia, 2017. Demographics of Russia --- Wikipedia, The Free Encyclopedia.
Wong, K.I., Wong, S.C., Yang, H., Wu, J.H., 2008. Modeling urban taxi services with multiple user classes and vehicle modes. Transp. Res. Part B Methodol. 42, 985–1007. doi:10.1016/j.trb.2008.03.004
World Meteorological organization, 2015. WMO OSCAR database [WWW Document]. URL http://www.wmo-sat.info/oscar/ (accessed 6.29.15).
Wu, C., Gerla, M., Mastronarde, N., 2015. Incentive driven LTE content distribution in VANETs. IEEE, pp. 1–8. doi:10.1109/MedHocNet.2015.7173168

Yang, H., Lau, Y.W., Wong, S.C., Lo, H.K., 2000. A macroscopic taxi model for passenger demand, taxi utilization and level of services. Transportation 27, 317–340.

Yang, H., Wong, S.C., 1998. A network model of urban taxi services. Transp. Res. Part B Methodol. 32, 235–246.

Yu, C.-H., Doppler, K., Ribeiro, C.B., Tirkkonen, O., 2011. Resource sharing optimization for device-to-device communication underlaying cellular networks. Wirel. Commun. IEEE Trans. On 10, 2752–2763.

Zegras, C., Litman, T., 1997. An analysis of the full costs and impacts of transportation in Santiago de Chile. International Institute for Energy Conservation.

Zervas, G., Proserpio, D., Byers, J., 2016. The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. Boston U Sch. Manag. Res. Pap.

Zha, L., Yin, Y., Yang, H., 2016. Economic analysis of ride-sourcing markets. Transp. Res. Part C Emerg. Technol. 71, 249–266. doi:10.1016/j.trc.2016.07.010

Zhang, W., Ukkusuri, S.V., 2016. Optimal Fleet Size and Fare Setting in Emerging Taxi Markets with Stochastic Demand. Comput.-Aided Civ. Infrastruct. Eng. 31, 647–660. doi:10.1111/mice.12203

Annex I Other federations

Taking into account the definitions here, some SoS that qualify as an FoS are the GEOSS (Lautenbacher, 2006), and, the EU Copernicus program (Copernicus Space Component Data Access (CSCDA), n.d.) as it includes third-party missions. Other cases which we can identify as federations are Vehicle to Vehicle (V2V) information networks (Festag, 2014), and Vehicle-to-Grid (V2G) energy exchanges (Liu et al., 2013). For completion we include here a brief introduction to the latter.

I.I Vehicle to Vehicle (V2V) cooperation

V2V communications, realized through VANET (Vehicular Ad-hoc Networks) is a proposal on the wake of the IoT and *Intelligent transportation systems* concept (Dimitrakopoulos and Demestichas, 2010). The value proposition on road vehicles 'talking' to each other (and also to the infrastructure) is based upon safety warning, road and traffic conditions information exchange, and additional services such as on-street unregulated parking finding. As such, V2V is a supporting technology of the self-driving car (Gerla et al., 2014).

In a VANET, cars exchange location, speed and acceleration to avoid collisions and evaluate the road status. The cars can also share data-centric resources in a very similar fashion to FSS, as shown in Figure 86. Moreover, some authors (Wu et al., 2015) propose use VANET for 3G/LTE mobile connection sharing, thus distributing internet content (for instance, for on-board entertainment) without needing to use a LTE mobile connection in each car. There are several standards and initiatives supported by car manufacturers and governmental institutions to make V2V a reality (Car 2 Car communication Consortium, 2011).

We can consider the V2V communications as the backbone of a federation of systems, that is, cars. The principles of goals, operational and managerial independence are respected. The cars cooperate to enhance their own operations. But, what are the costs incurred by V2V cooperation on each car?

Obviously the inclusion of some equipment in the vehicle, including location and communication devices is required, but this appears to be a very small fraction of the cost of manufacturing a road vehicle. There are also some hidden costs in terms of coordination, maintenance and development of standards by the car manufacturers, that will be added to the cost of V2V communications per car, but this is probably unnoticeable on the cost of an individual car due the large economies of scale car production enjoys. Hence, the inclusion of the V2V interfaces will hardly pose a design tradeoff if we only think in terms of manufacturing and deployment costs.

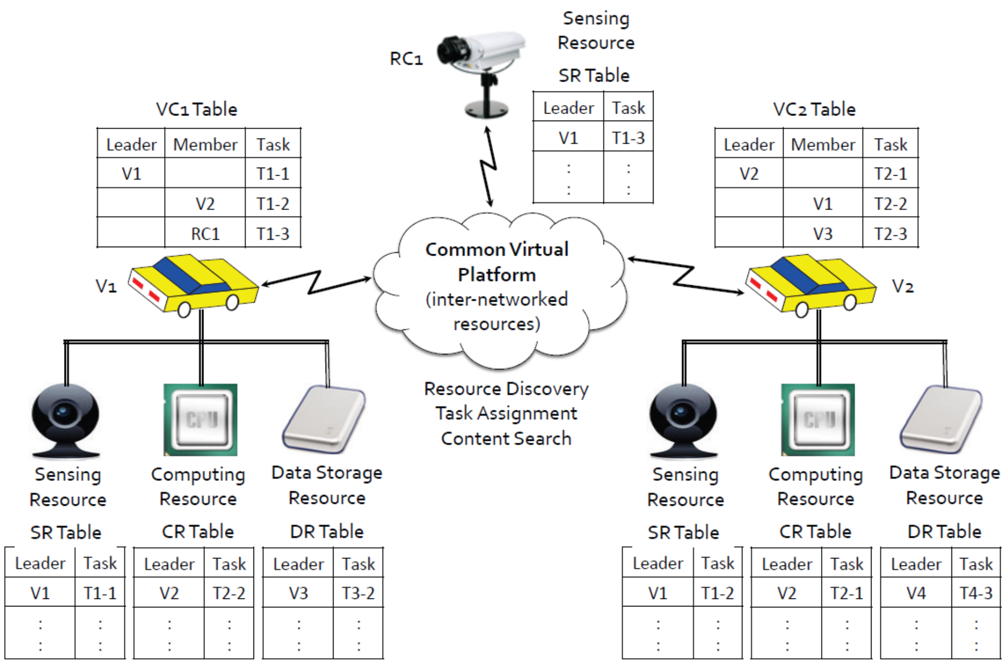

Figure 86. The cloud computing model applied to VANETs, from (Gerla et al., 2014).

However, there are other costs, not directly economical, of including V2V interfaces in a car, which concern the congestion of the radio spectrum, the potential security and safety concerns posed by network attacks, and the user privacy protection laws. We can conclude V2V is indeed an FoS, whose adoption largely depends on the regulatory environment rather than explicit cost-utility considerations by the cars' user. Figure 86 illustrates V2V.

I.II Vehicle to Grid

Proponents of Vehicle to Grid (V2G) and Grid to Vehicle (G2V) (Liu et al., 2013) recognize the opportunity for Electric vehicle (EV) to act as controllable loads and distributed sources in the smartgrid. EVs can act as capacitors if bi-directional chargers are in place, thus helping the grid overcome the frequency, voltage and reactive power problems posed by the increasing usage of renewable sources of intermittent nature.

In this case, the EVs, home grids, and the overall distribution grid compose a federation, which can exchange energy strategically to stabilize the grid and ultimately reduce power losses. Figure 87 illustrates this concept. As discussed, in this FoS, the main operand of the cooperation is electric energy, instead of data. The adoption dynamics in this FoS depend largely on the battle for the technical standards in car chargers, the specifics of the charging stations available, and the dynamics of every regions' grid, which might be able or not to handle bi-directional car chargers.

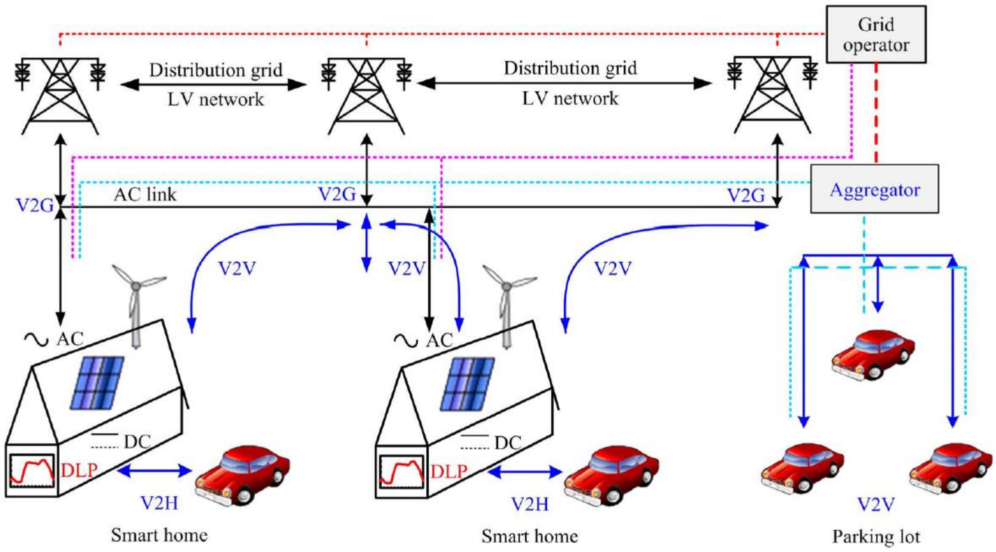

Figure 87. The grid concept featuring home, vehicle and distribution grids, and the different exchanges possible, from (Liu et al., 2013).

Annex II Notes on cost models

II.I Dedicated ground stations cost model

The cost model for dedicated ground stations is based upon COCOMO81 as in (Wertz and Larson, 1999) and (Wertz et al., 2011). Following the model, the development costs GS_{dev} are estimated as fractions on top of the software development, as Eq. II.I shows.

$$GS_{dev} = Sw + Fac + Log + Eqp \qquad \text{Eq. II.I}$$

Following the recommendations in (Wertz and Larson, 1999) the costs of facilities *Fac*, Equipment *Eq*p and Logistics *Log* are all fractions of the software cost *Sw* as in Table 51.

Table 51. Ground station costs as fraction of software costs

	Development cost as % of software cost
Facilities	18%
Ground equipment	81%
Logistics	15%

We assume here the first ground station development costs follows Eq.II.I and in any additional GS the software is re-used, cutting down the cost of recurring ground stations to the sum *Fac + Log + Eqp.*

The software cost *Sw* depends on the lines of code involved, new and re-used ($KLOC_{new}$ and $KLOC_{re}$), the effort adjustment factor *EAF*, the re-implementation factor *ES* and the monthly costs of software engineers *FTE*, as per the expression in Eq.II.II.

$$Sw = FTE \cdot (EAF \cdot 3.312 \cdot (KLOC_{re} \cdot ES + KLOC_{new})^{1.2}) \qquad \text{Eq.II.II}$$

Table 52 summarizes the values taken for the parameters above, following the recommendations in (Wertz et al., 2011).

Table 52. Model parameters for ground station software cost estimation.

Parameter	Symbol	Value adopted
Software engineer costs (experienced)	FTE	12,500 USD/month
Effort adjustment factor	EAF	0.91
New lines of code	$KLOC_{new}$	50K
Re-use lines of code	$KLOC_{re}$	100K
Re-implementation factor	ES	0.3

With the factors above, a first dedicated ground station costs about 15.5 MUSD and an additional one approximately 8.2 MUSD. In the FSS case study, we round these values to 15 MUSD for one station and 25 MUSD combined development cost for 2 ground stations.

The ground station maintenance costs are assumed to be 300 kUSD/yr per ground station, adding up the recommendations in (Wertz et al., 2011) for software and hardware maintenance. Other cost contributions such as missions operations are not counted since the cost assignment in the FSS case study is comparative and mission operations costs are incurred regardless of the mission communications architecture.

II.II Space-to-Ground link cost assumptions

Real data to evaluate the cost of the space segment hardware is hardly ever publicly available. Notably, some small satellite providers publish some of its hardware price lists providing a degree of validation of the orders of magnitude involved. In order to provide estimates for the mission data downlink hardware cost, we use here vendor data, the USCM-8 model, and several best-effort assumptions.

Vendors such as Syrlinks and SSTL (Surrey Satellite Technology US, 2017) provide 50 and up to 150 Mbit/s, very compact payloads with reduced mass and power footprint, tailored to small satellites. Given the pricings they propose for components we assume here 0.5 MUSD for the 50 Mbit/s performance (combining purchase of antenna, amplifier and transponder) and 2 MUSD for 150 Mbit/s.

From that point on we assume heavier solutions are needed, for which we can start to use the USCM8 model which has a range of validity starting at 39 kilograms. We will assume the data downlink subsystem hardware is not developed on purpose and follows from some heritage. We will then use USCM8's recurring unit cost for communications payloads (Wertz et al., 2011). Considering the mission data downlink as a communications payload might overestimate its cost as communications payloads typically operate with extended duty-cycles and can include beam-forming, electronic steering and other complex solutions (Maral, 2009).

To create a cost point for the 300 Mbit/s performance, consider Thales' Mission Payload Data Handling and Transmission system (PHDT) deployed in Cosmo-Skymed and other platforms (L'Abbate et al., 2014). The PHDT weights about 88 kilograms and provides 310 Mbit/s. USCM-8 estimates such equipment to cost about 17 MUSD. We adopt 10 MUSD instead to avoid over-estimating as argued above. This is a conservative assumption towards federation as it might reduce the cost of achieving high performances by an isolated spacecraft.

Finally, on the other side of the spectrum, the COCOMO81 proposes 30 MUSD for a 160 kilograms payload. We link the latter to the highest datarate option (500 Mbit/s). This is approximately as doubling the PHDT capacity. In this case we do not trim down the cost as such performance is near the state of the art of mission data downlink equipment (Rosello et al., 2012).