PhD Thesis Work Plan

SITI Technical Report SITI-TR-11-16

2011

Authors/Editors

Author	Affiliation
Bruno Batista	Universidade Minho/SITI, University Lusófona

UNIVERSIDADE LUSÓFONA de Humanidades e Tecnologias

Tel: 217 515 500 | Fax: 21 757 7006 | http://siti.ulusofona.pt/

SITI, Universidade Lusófona
Campo Grande 376, 1749-024 Lisboa
Portugal

DOCTORAL PROGRAMME IN COMPUTER SCIENCE

PhD Thesis Work Plan

Information Centric Opportunistic Networks

Bruno José Duro Batista

September 2011

Advisors: Prof. Dr. Paulo Milheiro Mendes (University Lusófona), and Prof. Dr. Paulo Mateus Mendes (University of Minho)

siti
informatics systems and technologies research unit

Contents

Abstract

The Internet is a success due to its clear utility to regular people. At the same time, the evolution of mobile communications and devices allows to get the capability of sensing the world, store information and exchange it between us in an opportunistic way. However, the Internet end-to-end communication model based upon the location of devices, is not in line with users' solely interest in information and not on its location. This work aims to investigate a solution to mitigate the mismatch between network architecture and users' interest on information by replacing the end-to-end Internet paradigm by an information-centric paradigm, in which networking is based on information itself and not upon the location of end-devices. To allow users to access/exchange information at any time, the new networking architecture aims to be able to operate even in the presence of intermittent connectivity.

Chapter 1

Introduction

Nowadays a large amount of people uses the Internet for some kind of activity (business or pleasure) on a daily basis. The Internet moved from the bulky desktop computer to more pervasive devices like our smart phones or some form of embedded device that, for instance, can be found in the increasingly popular intelligent buildings, in our homes (in fridges, laundry machines, domotic systems, etc), on the street (for monitoring purposes, advertising, etc). The increase capabilities in terms of CPU, storage and communication capabilities of these devices means for one hand, that they can be consumers and producers of information, and on the other hand, exacerbates, among others, problems related with mobility and opportunistic networks (in opportunistic networks users take advantage of the diverse communication, computation, sensing, storage and other resources that surround them to produce and share information).

Looking to when and how the people uses the Internet (and network systems in general) a clear mismatch can be found. People only care about obtain information, they do not care where the information is. From the network point of view the location is essential to find the requested content, thus when a person asks for a content by name, that name has to be resolved to a location and a name that the network can identify. The current network architecture, very dependent of content and hosts location, is far from ideal in scenarios where hosts are very mobile, have a multitude of communication interfaces (with very different technologies) that need to work in a seamless way and have intermittent connectivity like ad-hoc and opportunistic networks.

This work aims to investigate a solution to mitigate the mismatch between network architecture and users' interest on information based on the information-centric paradigm applied to communication networks, made available by the action of personal devices. The general guideline is to rethink the alignment of data communications and routing with the fundamentally broadcast nature of intermittently available wireless channels. To analyze the most suitable information-centric networking approach for opportunistic networks, where connectivity is intermittently available, this work will investigate the correlation between routing and name resolution and in-network storage for caching. The integration of caches inside a network, and how such caches should be synchronized in terms of available data, will be investigated, as well as the role of the naming resolution system (e.g. based on flat or hierarchical naming schemes) in the routing of information among data containers. From the architecture point of view the relationship of networking architecture innovation versus the so-called over the top approaches in the application layer, such as Haggle, will be investigated. In order to proof the deployment factor of the proposed information-centric opportunistic networking approach, this work aims to analyze the support that information-centric networking can give to an Internet of Things. This means that use-cases and test-beds based on personal sensing devices, such as smart phones and embedded devices, will be used since this type of devices are pervasively available everywhere, and they already incorporate (or have capability to incorporate) sensorial equipment such as 3-axis accelerometer, gyroscope, proximity sensor and GPS. Personal computing devices have a huge capability to generate information (environmental and personal), being a suitable platform to test an information-centric architecture.

This document is structured as following. Chapter 2 presents the most relevant related work focusing on different solutions that have been proposed in what concerns overall information-centric architectures and solution to core characteristics (routing, name resolution and placement of in-network caches). In Chapter 3, all the contributions expected from this thesis work are listed. Chapter 4 shows the experimental methods and tools considered to be used throughout the development of this work, whereas Chapter 5 discusses the main activities along with the expected outcomes and the methodology to be employed in order to achieve them. And, finally, Chapter 6 shows the major scientific milestones expected with the development of this thesis.

Chapter 2

State of the Art

This section starts by providing an overview of several proposals aiming to devise a data-centric networking architecture that, one way or another, address issues related to the information centric paradigm and opportunistic networks. The number and the funds involved in development of these proposals, as well as the set of Internet experts, some of them were already in the genesis of the current Internet architecture, are a clear sign of the great interest of the research community in topics related with this new way of looking networking. The goal of analyzing such proposals is twofold: i) to identify the most significant trends aiming to investigate an information-centric networking paradigm; ii) to identify the potential impact that such architectures may have in more challenging scenarios encompassing a significant probability of intermittent connectivity.

The analysis of the presented set of architectures allows the identification of several technical challenges, from which two are selected as the target of this thesis investigation: routing and name resolution, as well as in-network caching in self organized systems.

The analysis of the state of the art in the two identified technical areas shows that models based on an information-centric paradigm present themselves a potential solution for the new challenges posing to networking (especially in disruptive environments), both in content delivery and real time applications [10], while trying to reduce the gap between the way the current Internet architecture works (based on hosts) and the use that people wants to make of it: access to information while being agnostic about its location.

However the majority of the analyzed architectures and the routing, naming and in-networking caching solutions have in mind the Internet backbone and (wireless/wired) local access networks, aiming to devise a future Internet architecture. From the best of our knowledge, there is few prior work that address completely the problem of analyzing the most suitable information-centric networking approach for opportunistic networks: networks in which users take advantage of the diverse communication, computation, sensing, storage and other resources that surround them to produce and share information.

2.1 Data Centric Architectures

2.1.1 General Description

4WARD Project

The EU funded 4WARD (http://www.4ward-project.eu/) project (January 2008 to June 2010) aimed to increase the competitiveness of the European networking industry and to improve the quality of life for European citizens by creating a family of dependable and interoperable networks providing direct and ubiquitous access to information. To fulfill this objective 4WARD partners worked to overcoming the shortcomings of current communication networks like the Internet in a framework that allows the coexistence, inter-operability, and complementarity of several network architectures in an integrated fashion, avoiding pitfalls like the current Internet's "patch on a patch" approach.

The 4WARD project developed the NetInf (network of information) architecture based on information centric paradigm and implemented a prototype the OpenNetInf (available at http://www.netinf.org/home/home/). The architecture of NetInf considers the existence of NetInf nodes (NIN) which provide an API, that supports searching for information and resolving information IDs into appropriate content locators. This nodes are able to perform lookup and search locally. This architecture also considers the existence of an infrastructure of NetInf nodes that provides these services on a global scale. These services are used by other nodes (clients) to easily implement various information-centric applications. In NetInf all the information is stored in Information Objects (IO's) using a flat naming scheme. The infrastructure consist in Identifier Lookup Service (ILS) and the IO Lookup Service (IOLS) for name resolving and searching purposes. To use the infrastructure, client applications communicate with the IOLS and ILS via the NetInf API. This API is provided by the NetInf Node Middleware (NNM) [6].

Sail Project

The EU funded Scalable & Adaptive Internet Solutions (SAIL) project (http://www.sail-project.eu/) started in August 1st, 2010. It´s objective is the research and development of novel networking technologies using proof-of-concept prototypes to lead the way from current networks to the Network of the Future.

SAIL continues the work started in the 4WRD project in the investigation of the NetInf architecture aiming to improve application support via an information-centric paradigm, replacing the old host-centric one. This project enables the co-existence of legacy and new networks via virtualisation of resources and self-management, fully integrating networking with cloud computing to produce Cloud Networking (CloNe). It also embraces heterogeneous media from fiber backbones to wireless access networks, developing new signaling and control interfaces, able to control multiple technologies across multiple aggregation stages, implementing Open Connectivity Services (OConS). SAIL furthermore addresses several topics which 4WARD did not or only partly covered. Examples are content transport, interdomain issues, congestion control and cache management, all of them targeting a new Internet backbone architecture.

Pursuit Project

The EU funded Pursuit project (http://www.fp7-pursuit.eu/PursuitWeb/) started in September 2010 as a continuation of the also EU funded PSIRP project. The core goals of the Pursuit project are twofold: i) develop information-centric inter-networking solutions in crucial areas like transport, caching, error control, flow control; ii) investigate new techniques for building wireless and wireline networks and access methods applying information centrism throughout the layers. The devised solution is to be evaluated based on relevant quantitative and qualitative parameters , with a focus on security and privacy, tested in integrated prototypes and demonstrators that show the full potential of the developed solutions. Moreover the Pursuit project aims to investigate the socio-economic impact of the developed solution both on macro-economic and player-specific level.

The Pursuit approach, as its precedent PSIRP, follows a concept of information centric based on a publish-subscribe paradigm. In Pursuit Information items are organized within scopes. Scopes may denote physical structures, e.g., a corporate network, or logical structures, e.g., my friends in facebook. Every information item is uniquely identified, using a flat naming scheme, by a pair of identifiers; a rendezvous identifier (RId) and a scope identifier (SId). In order for a publisher to issue a publication it has to know the SId of a scope within which it wants the publication to be published, as well as to create a RId for the publication. The publication's RId is then forwarded to the rendezvous node of the SId rendezvous network, which manages the publication's RId. Pursuit's operation is based on three basic functions: the rendezvous, the topology and routing, and the forwarding functions. The rendezvous function is responsible for matching subscribers' interests with publications. The topology function monitors the network topology and the routing function is responsible for creating information delivery paths. Finally, the forwarding function implements information forwarding throughout the delivery paths.

Named Data Networking Project

The Named Data Networking (NDN) project (http://www.named-data.net/; http://www.ccnx.org/) started in November, 2010 and is funded by the National Science Foundation (USA) under the Future Internet Architecture program. NDN follows the work started by Van Jacobson (were he is the solution architect) at the Palo Alto Research Center. NDN tries to capitalize on strengths, and addresses weaknesses, of the Internet's current host-based, point-to-point communication architecture in order to naturally accommodate emerging patterns of communication. By naming data instead of its location, NDN transforms data into a first-class entity. The current Internet secures the data container (communication session) while NDN secures the content itself, a design choice that decouples trust in data from trust in hosts, enabling several radically scalable communication mechanisms such as automatic caching to optimize bandwidth. The project studies the technical challenges that must be addressed to validate NDN as a future Internet architecture: routing scalability, fast forwarding, trust models, network security, content protection and privacy, and fundamental communication theory. The project uses end-to-end testbed deployments, simulation, and theoretical analysis to evaluate the proposed architecture, and is developing specifications and prototype implementations of NDN protocols and applications.

Associated to the NDN project some initial studies of the application of information-centric architectures to multihop wireless networks were performed [17]. While the research findings show the alignment of the information-centric paradigm with the more dynamic behavior of multihop wireless networks, several challenges still remain to handle scenarios were end-to-end communication between any two pair of nodes may not exist all the time, e.g. how to use the cache mechanisms to improve data delivery and how to deal with source and destination mobility.

Connect project

The Connect project (http://www.anr-connect.org/), funded by the Agence Nationale de la Recherche (France) and started in December 2010, aims to contribute for the definition and evaluation of a new paradigm for the future Internet: a content-centric network (CCN) where, rather than interconnecting remote hosts like IP, the network directly manages the information objects that users publish, retrieve and exchange. The project is built on existing CCN proposals, adopting as a starting point the concept currently promoted by the Named Data Networking project. The Connect project tries to complement the existing work on CCN in the following three technical areas: Traffic control; Naming, routing and forwarding; Replication and caching strategies. Regarding traffic control Connect will investigate new, name-based criteria to ensure fairness and to realize service differentiation, since the authors consider that TCP queue management is not adequate anymore. As for naming, routing and forwarding they consider that CCN choices are often expedients to facilitate overlay implementation, so the intent is to prove that the name-based routing and forwarding is scalable and to design algorithms suitable for full-scale implementation. Finally concerning replication and caching strategies they will define replication and caching strategies and evaluate their performance.

Comet Project

The COntent Mediator architecture for content-aware nETworks (Comet) project (http://www.comet-project.org/) started in January 2010 and is funded by the European Commission's 7th Framework Program. Comet aims to provide a unified interface for content access whatever the content characteristics are: temporal nature (pre-recorded or live), physical location (centralized or distributed), interactivity requirements (elastic or real-time). It also aims to apply the most appropriate end-to-end transport strategy for content delivery, by mapping the content according to its requirements and user preferences to the appropriate network resources (best quality of experience (QoE) for end users) and to support all different types of content distribution (Unicast, Anycast, Multicast, P2P-delegated).

The Comet approach introduces a content mediation plane that combines content resolution and access: locating content according to delivery requirements (content mediation) and delivering it using the most suitable resources (network mediation). The architecture addresses two key aspects: Global content naming and addressing, to support content discovery and name resolution to the identifiers required for an unified access; Network-aware distribution by mapping the content onto the appropriate network resources, providing content and network-aware access to every type of content through all possible types of distribution.

Haggle Project

The EU Haggle project (http://www.haggleproject.org/) (2006 - 2010) proposed a networking architecture for content-centric opportunistic communication. The main objective was to develop a new autonomic networking architecture, designed to enable communication in the presence of intermittent network connectivity, by exploiting autonomic and opportunistic communications (i.e. in the absence of end-to-end communication infrastructure). The Haggle proposal departs from the existing TCP/IP protocol suite, completely eliminating layering above the data-link, and exploiting application-driven message forwarding, instead of delegating this responsibility to the network layer. To this end, the project went beyond already innovative cross-layer approaches, defining a system that uses real best-effort, context aware message forwarding between ubiquitous mobile devices, in order to provide services when connectivity is local and intermittent. Only functions that are absolutely necessary and common to all services are used. The goal was to provide a more human-oriented way of communicating, rather than deploying mechanisms that are related to the technological aspect of the communication. Haggle implements a data-centric communication model with a publish-subscribe (pub/sub) API, which spreads application data from device to device based on the data's match against a user's declared interests [21].

2.1.2 General Conclusions

From the presented information centric architectures CCN and Haggle seams to be the most suitable candidates to be the starting point for our work and to improve upon.

From the architectures targeting the Internet backbone and access networks the initial findings lead us to believe that CCN is possibly the most suitable because, as NetInf, has a flexible in-network caching system, which can work without any kind of central coordination, and thus can support the functionality of self-organization that will be developed in this work. Also CCN, on contrary to NetInf and the other architectures, supports node mobility implicitly, without the need of rendezvous points or any other intermediary that can jeopardize the desirable seamless operation. In CCN a client can switch network and continues to issue interest packets (and retransmit if needed). Content provider mobility is more difficult in all architectures and needs to be addressed with care.

Regarding architectures targeting opportunistic networks Haggle is the only one that was developed having in mind the opportunistic networks scenario and it has already incorporated some concepts of the information centric paradigm. Despite that Haggle focus was mainly develop a flexible architecture that allowed the exchange of information in opportunistic networks. It's approach is fundamentally different from CCN since Haggle considers a layerless network model, in order to make it more flexible and reduce feature interaction problems [3].

Considering this, the plan is to start with the concepts developed in Haggle regarding opportunistic networks and adapt/enhance them to complement the CCN model.

2.2 Challenges for Opportunistic Networks

2.2.1 Routing and Name Resolution Approaches

In light of the information-centric paradigm, the location of the information is not relevant, the fundamental is the content itself. This fact raises several questions related with almost every networking technology (it is a complete shift of how networking is done) e.g. how to name objects, how to route and forward this information, how to guarantee security, how ensure traffic control and delivery, how to manage in-network cache, how to do mobility management, etc. Among all these questions two of them are of particular importance, due to the impact in the networking process: how to name information, where to stored it and for how long and how to forward it. Decoupling from location means that every content available needs to have a unique name and that it must be routable by name.

2.2.1.1 Internet backbone and access networks

To achieve a content-centric architecture, T. Koponen et al. [12] propose to replace DNS names with flat, self-certifying names, and replacing DNS name resolution with a name-based anycast primitive that lives above the IP layer. They called it the Data-Oriented Network Architecture (DONA). The focus of this work is on naming structure and name resolution. The name structure proposed by DONA is organized around principals. Each principal is associated with a public-private key pair, and each datum or service, or any other named entity (host, domain, etc.), is associated with a principal. Names are of the form P:L where P is the cryptographic hash of the principal's public key and L is a label chosen by the principal, who ensures that these names are unique. When a client asks for a piece of data with name P:L and receives the triplet <data, public key, signature>, it can immediately verify that the data did indeed come from the principal by checking that the public key hashes to P, and that this key also generated the signature. DONA uses the route-by-name paradigm for name resolution. As a consequence, rather than using DNS servers, DONA relies on a new class of network entities called Resolution Handlers (RHs). Name resolution is accomplished through the use of two basic primitives: FIND(P:L) and REGISTER(P:L). A client issues a FIND(P:L) packet to locate the object named P:L, and RHs route this request towards a nearby copy. REGISTER messages set up the state necessary for the RHs to route FINDs effectively. Authors had implemented a Linux-based prototype of an RH and deployed it on Planetlab. The implementation is useful to validate the concept but is not of sufficient size (nor subject to sufficient workload) to allow the study of DONA's scaling properties. Nevertheless, authors did not prove the scalability properties of the model. Moreover, the model was developed considering the Internet scenario with good connectivity and did not support, natively, more dynamic scenarios such as opportunistic networks.

Van Jacobson's approach, called Content Centric Networking (CCN) [11], is to build upon successful features of TCP/IP, with the key change of replacing the machine-oriented IP model with a named content model as the basis for the central protocol that connects networks. Using new approaches for routing named content, similar to IP routing — and including integrated content-based security — authors propose to achieve scalability, security, and performance in a single unified solution to myriad networking problems. CCN differs from IP networking in several ways, namely i) CCN can take maximum advantage of multiple simultaneous connections due to its simpler relationship with layer 2; ii) CCN ensures the security of content itself, rather than the connections over which it travels, thereby avoiding many of the host-based vulnerabilities that plague IP networking. CCN model defines two packet types, Interest and Data. A consumer asks for content by broadcasting an Interest packet over all available connectivity. Any node hearing the interest and having data that satisfies it can respond with a Data packet. Data is transmitted only in response to an Interest and consumes that Interest. Because CCN's forwarding model is a strict superset of the IP model with fewer restrictions (no restriction on multi-source, multi-destination to avoid looping) and has the same semantics relevant to routing (hierarchical name aggregation with longest-match lookup), any routing scheme that works well in IP should also works for CCN. Nevertheless, the CCN model is created and validated considering networks with good connectivity all the time, and does not provide a good insight about how to handle the number of copies of information cached in the networks in order to reduce delays and increase robustness.

In contrast to CCN, NetInf naming scheme is based on flat names. In [7], authors presented the NetInf naming scheme that has been developed for flexibility and extensibility by supporting multiple types of IDs. The naming scheme relies on proved mechanisms like cryptographic hashing and public-key certificate chains to reduce the risk of vulnerabilities. Authors had identify the desired characteristics of a naming scheme: be generic so that it can name any kind of entity, including static and dynamic Information Objects (IOs), services, network nodes, people, and real world entities like places and objects; content must be self-certified, self-certification ensures the integrity of data and securely binds this data to an ID; names should be persistent, not only with respect to storage location changes, but also with respect to changes of owner and/or owner's organizational structure and content changes producing a new version of the information. In order to achieve trust and accountability authors present two mechanisms: owner authentication, where the owner is recognized as the same entity (as the object owner), but may remain anonymous, and owner identification, where the owner is also identified by a physically verifiable identifier, such as a personal name. NetInf was developed to meat these requirements and thus securely name a wide variety of objects and IOs. To support this variety and potentially diverse requirements, the NetInf naming scheme enables flexibility and extensibility by supporting different name structures, differentiated via the Type tag in the identifier. In the NetInf naming scheme, any entity/IO is represented by a globally unique ID. Together with the IO's data and metadata, an IO is defined as IO = (ID, Data, Metadata). Data contains the main information content of the IO. Metadata contains information needed for the security functions of the NetInf naming scheme, e.g., public keys, content hashes, certificates, and a data signature authenticating the content. It also includes non-security-related information, i.e., any attributes associated with the IO, e.g., the location where a picture was taken. Metadata can be an integral part of the IO or can be treated and stored independently, as a separate IO. Authors had evaluated the presented NetInf naming scheme in a Java based prototype [6] and proved its feasibility considering two applications: an email client (thunderbird) and a web browser (firefox). Despite that authors did not prove the scalability, which is a great concern especially when using flat names, of the NetInf naming scheme and the test scenario was limited where authors considered very specific applications and did not test in dynamic conditions.

2.2.1.2 Dynamic networks

Aiming to target more dynamic networks, S. Oh et al. proposed a Content Centric Networking (CCN) protocol suite [22] that supports content addressing, repository, and distribution in a large wireless Mobile Ad hoc NETwork (MANET), specially in tactical/emergency networks. The most important feature of such networks is group-based mobility. Unlike Internet, content in the tactical/emergency MANET can be classified as topic based content and spatial/temporal content. Local storage is a key function of the CCN and the MANET CCN also supports it. Tactical MANET CCN employs two kinds of routing schemes: content pushing and pulling. On the other hand, Internet CCN uses content pulling and content announcement pushing. Due to wireless shared medium, the MANET CCN has more features than Internet CCN: employs interest aggregation and has packet collision avoidance mechanism employing control packets, Reply and Request. CCN supports content-based security as required in tactical and emergency communications: using PKI, all transmitted packets are authenticated with digital signatures or encrypted by a public key. Despite the fact that authors consider the in-network cache functionality of CCN to be very important they assume that handheld devices only have a small amount of storage capacity and so they only save shortly the received content in the buffer. Data files are stored when possible using LRU or LFU replacement policy and they only can be transmitted to direct neighbors. The hierarchical nature of the network means that the content repositories, which contains the spatial/temporal information and the data itself, are only available in the backbone network. The MANET CCN model considers a very limited scenario (tactical and emergency MANET with a hierarchical network structure), and the evaluation is done in a small test-bed, that rise some questions about scalability.

Another example of an application of CCN model in a dynamic scenario is the work proposed by Michael Meisel et al. [17]. Authors start by elaborate about the limitations that current network model poses in scenarios where nodes are very mobile and how and why CCN model deal with this problems. They advocate that the problems related with node mobility arise from the fact that node need IP to communicate. Since the IP is direct link to location when the node moves the IP needs to change and the connectivity can be lost (at least partially). Since CCN decouple information from location these problems cease to exist, which is a good indication of the suitability of the CCN model to dynamic networks. Considering this, authors developed a new forwarding algorithm Listen First, Broadcast Later (LFBL) [18], designed from the ground up with named data in mind. LFBL uses a variation of NDN's 3-way exchange (name pre announcements, Interest forwarding, and data return) meant for a mobile, ad-hoc environment. A node N floods a request that carries the name of the requested application data. Any node or nodes that happen to have that named data send a response packet, which is forwarded to N using information gleaned by the intermediate nodes during the request phase. Finally, N sends an acknowledgment as feedback to encourage or discourage these responses. To evaluate the performance of LFBL, authors implemented LFBL in the QualNet network simulator and compared the performance against AODV. According to they, LFBL out-performs AODV using four different metrics and in a diverse set of simulation scenarios. The results shown that, under high dynamics, LFBL delivers nearly 5 times more data compared to AODV, while having comparable overhead, introduced mainly due to redundant packets. Despite working over dynamic networks, authors assumed the existence of at least one end-to-end path among any pair of nodes at any time, which may not happen all the times due to absence of direct links in sparse networks, or intermittent connectivity due to interference in dense networks.

Since the research related with the information centric paradigm is, for the most part, in its early stages, the considered scenarios in previous research, as shown above, are somehow limited to wired networks and very specific wireless networks. Considering the scenario of opportunistic networks, there are not any work that evaluates the full potential of information centric concepts in this king of scenarios. The work that is closest was developed under the Haggle project [24]. Haggle presents itself as novel framework that enables seamless network connectivity and application functionality in dynamic mobile environments. Haggle allows separating application logic from the underlying networking technology. Applications should not have to concern themselves with the mechanisms of transporting data to the right place. Instead, this should be left to the networking architecture. To achieve this separation, Haggle uses a data-centric network architecture that internally manage the task of handling and propagating data. Thus, applications can be automatically adapted to dynamic network environments using the best connectivity channel for the situation. Haggle is internally composed by six managers organized in a layerless fashion: name manager; data manager; resource manager; connectivity manager; protocol manager; forwarding manager. Each manager is responsible for a key modular component or data structure. This provides flexibility, e.g. allowing for many protocols (e.g. SMTP and HTTP) and connectivities (e.g. 802.11 and Bluetooth) to be instantiated simultaneously. The managers and modules have well-defined APIs, and each manager (and internal component) may use the API of any or all of the other managers. The authors state that this architecture provides necessary and useful flexibility over a layered architecture in which each layer may only talk to the two APIs above and below. In Haggle, connectivity is regarded as a schedulable resource which can consume network time, battery power, monetary costs, etc. Haggle encapsulates the late-binding of communication protocols and forwarding algorithms necessary for transporting data. Communication protocols specify the method for point-to-point communication, both for transmitting data as well as opening and receiving connections. Haggle's architecture allows many forwarding algorithms to be in use simultaneously. To be able to forward data, the authors propose a general form of naming notation that allows late-binding of many user-level names, independent of the lower-level addressable name. They achieve this by using name graphs, which are hierarchical descriptions of many known mappings from a user-level endpoint to lower-level names. In the Data Manager Haggle exports an interface for applications to manage persistent data and metadata explicitly. Haggle's data format is designed around the need to be structured and searchable. The authors demonstrate the effectiveness of Haggle's approach using existing email and web applications on a Haggle prototype. The tests show the ability to dynamically select the best network operating mode when transferring emails and function, even when disconnected from infrastructure. Despite the fact that Haggle incorporates concepts related with information centric the focus is on sharing information, with mutual interest, in a opportunistic way. The developed architecture is built, as far as possible, on top of the existing network technology, at application layer and thus do not incorporate, as desirable, the new concepts in the network itself (the decision of what information is forwarded and to which communication interface is taken by the application).

2.2.2 Self-organized pre-placement of in-network caches

One of the features of an information-centric architecture that is mentioned in all projects described in section 2.1, is to consider the existence of caches within the network. These caches can be implemented in different ways, e.g., it's possible to consider the existence of a cache in every nodes of the network or only in specific nodes, but its importance is unquestionable. Caches contribute to load balancing traffic and reduce perceived latencies, since the information comes from a "near" location, assist systems to be more fault tolerant and helps to reduce exploration and usage costs. In mobile environments, more specifically in opportunistic networks, the existence of caches is even more relevant since connectivity may be intermittent and the operation and usage costs need to be managed with caution. If two different persons, in the vicinity of each other, want the same content, it is counter-productive to both download it by 3G or GPRS. If possible, it's better for one to download it from the source and the other get it from the cached copy. This king of thinking it is not new and was already been applied in web caches, content delivery networks, etc. However it was not consider as part of the network model. Taking in account the scenario of opportunistic networks then we need to add, to the list of desired properties of the "cache system", the absence of any central coordination and the existence of some algorithm to detect communities that share some kind of interests (usage patterns for content objects). This means that the system needs to be self-organized in order to know what content to cache and where to cache it.

2.2.2.1 Self-organization

The concept of self-organization have been applied in several forms, with very different meanings and to solve many different problems [1]. In [13] authors proposed an adaptation of a self-organizing map (SOM) to community detection. A SOM is a type of artificial neural network that is trained using unsupervised learning. Network communities can be qualitatively described as graphs that represent the density of network nodes. A SOM represents topographic organizations in which nearby locations in the map represent inputs with similar properties. Authors validated the proposed algorithm using benchmark computer generated networks and several social based networks. The results shown the scalability of the proposed method, its accuracy (close to 100%) and that the algorithm can successfully identify communities independently of their size and heterogeneity. However, the number of communities must be defined, instead of emerged from the network itself, which means that its application in very dynamic networks can be problematic.

The communication among nodes in a self-organized system may be supported by a Peer-to-Peer system (P2P). However, P2P systems suffer with the dynamic behavior of nodes. Structured P2P systems may be difficult to administer in the case of high churn rate, because new or modified resources must be immediately (re)assigned to the corresponding peers. Recently the trend is to investigate "self-structured" P2P systems, so called because the association of descriptors to hosts is not pre-determined but adapts to the modification of the environment. A. Forestiero et al. [4] presented a P2P system, called Self-Chord, that inherits from Chord the ability to construct and maintain a structured ring of peers, but features enhanced functionalities achieved through the activity of ant-inspired mobile agents. In Self-Chord, the peers are organized in a logical ring. For each peer is given an index, having B_p bits, which is obtained with an uniform hash function and can have values between 0 and $2^{B_p} - 1$. Each resource is associated with a binary key, having B_c bits, which is used to discover and access to resource. The ring is constructed and maintained as in Chord, however, to inherit the efficiency of resource discovery operations (logarithmic) offered by it, the resource keys must be ordered on the ring. While in Chord, ordering is the outcome of a global planning, in Self-Chord it is obtained by the operations of ant-inspired agents that move the resource keys across the ring. This agents do not operate forever, but are generated and die as the ants to which they are inspired. Each peer of the ring, at time of its connection to the network, generates a mobile agent with a given probability P_{gen}. The lifetime of this agent is randomly generated with a statistical distribution whose average is related to the average connection time of the connecting peer and that is calculated on the past activity of this peer. Therefore, the turnover rate and the average number of operating agents are related to the dynamic characteristics of the network, i.e., to the frequency of peer joinings and departures. In order to validate self-chord, authors did several simulations with different distribution of keys. They also analyzed the behavior of self-chord under dynamic conditions and they did a scalability analysis. The paper also introduces an analytical framework that models the system evolution through a set of differential equations. However, the study is limited to P2P systems with a ring structure. Authors referred that it could be applied with other structures but the performance in this scenario is not validated yet. Moreover, the model should be validated in a real scenario.

Independently of the communication system among self-organized nodes, of major importance is the fault-tolerance property of an self-organized network in the sense that tasks can be shared in the network, and if some nodes break down then tasks can be relocated to other nodes. In [5], authors presented a Reconfigurable Self-Organized System (RSS), which is a modular system based on reconfigurable hardware technology, which is able to organize and restructure its own activity by interactions of its intelligent nodes without external influences. It has two main characteristics: the first one is the complete integration of the platform into a global environment and the second one is the autonomy of each node. As a networked system, the proposal is to integrate standard solutions for the inter-node decentralized communication. The communication functionality is divided into two parts: the first one is used for the task distribution using a request stream that is built to broadcast a request for helping other nodes; the second part is used to transfer data and bitstream files. The proposed approach has been designed and applied for real time video application in order to show the influence of a self-organized system in the existing technology. Nevertheless the study of the communication and the reconfiguration of a network and the reconfiguration of intelligent nodes are open questions.

Besides the detection of similarities in self-organized network, the inter-node communication system and the required fault-tolerance properties, it is important to emphasize the need to have a network in which all nodes have an accurate view of the overall system. To mitigate the potential lack of accuracy due to an ineffective acquisition phase, A. Tyrrell et al. [25] proposed to integrate synchronization into the communication phase. The proposed Meshed Emergent Firefly Synchronization (MEMFIS) model is based on the Mirollo and Strogatz [20] mathematical model for firefly synchronization (MS model). MEMFIS relies on the detection of a synchronization word common to all nodes, which is embedded into each packet along with payload data. The detected timings of received synchronization words serve as input to the slot synchronization algorithm, which is placed on the physical layer and based on the MS model. MEMFIS is not a synchronization protocol exchanging explicit timing information in the form of time stamps. From the Medium Access Control (MAC) layer point of view, slot synchronization emerges while nodes transmit data. Apart from the insertion of the synchronization word, which anyhow is required to decode payload data, no additional overhead or signaling between nodes is required. The performance of MEMFIS is evaluated against three key requirements: the scalability with respect to the network size, the adaptability with respect to different network topologies, and the robustness in case of unreliable detection of synchronization words. Authors have evaluated MEMFIS in a simulated environment and concluded that synchronization emerges faster and with low signaling overhead. Nevertheless, MEMFIS needs to be study and validated in more dynamic environment (like opportunistic networks). Although the proposed algorithm is very focused on wireless slot synchronization, the Mirollo and Strogatz model may be of interest when applied to allow the synchronization of the caches.

To have an accurate view of the overall system in a dynamic network, nodes must be able to discover in real-time the capacities of other nodes. In [4] authors presented a distributed grid information system supported by swarm intelligence for efficient resource discovery using flooding-like protocols. The proposed framework utilizes ant colony algorithms to build, optimize and maintain a self-structured peer-to-peer overlay network connecting grid nodes both using a minimal number of links, and ensuring that the diameter of this overlay network is bounded. To be able to execute resource discovery efficiently, the framework proactively query the overlay network with the purpose of locating nodes with similar capabilities and storing this information in a local cache for every node, so as to minimize the amount of queries being propagated throughout the network to find matching nodes. Authors defined resource discovery as the process of finding peers whose dynamic profiles (a profile can be viewed as a collection of tuples, expressed as a vector, referring to different resource aspects) match a given search query. From the simulations, the authors confirmed the merits of the proposed method for resource discovery, compared to traditional flooding methods and they concluded that the constructed topology, by the self-organized grid overlay, ensures bounded average path length with a minimal per-node degree and a minimal number of redundant links. Nevertheless authors only tested the framework in a single network topology, in which parameters considered are static.

2.2.2.2 Caching

The problem of placing caches closer to the end user it had already been subject of study although not in the same way that is consider in this work. In [2] S. Bhattacharjee et al. presented a study that highlighted the benefits of associating caches with switching nodes throughout the network, rather than in a few hand-chosen locations. Authors also considered the use of various self-organizing or active cache management strategies, in which nodes make globally consistent decisions about whether or not to cache an item in order to reduce overall latency. The conducted tests were simulated in AN-Sim and consider that the nodes able to cache content are uniformly distributed (they abstract the problem of select where to place the cache). The tests shown that active caching is beneficial across a range of network topologies and access patterns, and is especially effective when access patterns exhibit significant locality characteristics. In addition it was presented a simple analytic model for the expected one-way latency in accessing an item in a network that has intra-network caching. The basic model makes the following assumptions: the set of items is partitioned into "popular" items and "unpopular" items; only popular items are cached; every cache is full; the cached items encountered from one client access to the next are random and independent. The analytic models were simplified for tractability, though still provide reasonable estimates of access latency. Regardless the good results the test scenario was somehow limited to web usage, and only through simulations, in well connected networks and authors did not demonstrate the scalability and the overheat generated by the active solution.

Another possible approach is consider a cache system based on a peer-to-peer network. In [9] authors presented Squirrel a decentralized peer-to-peer web cache. Web caches are often deployed on dedicated machines at the boundary of corporate networks, and at Internet service providers. The key idea in Squirrel is to allow client desktop machines themselves cooperate in a peer-to-peer fashion to provide the functionality of a web cache. Currently, web browsers on every node maintain a local cache of web objects recently accessed by the browser. Squirrel enables these nodes to export their local caches to other nodes in the corporate network, thus synthesizing a large shared virtual web cache. That way, each node performs web browsing and web caching. Squirrel uses a self-organizing, peer-to-peer routing substrate called Pastry as its object location service, to identify and route to nodes that cache copies of a requested object [23]. Squirrel thus has the advantage of requiring almost no administration, compared to conventional cooperative caching schemes. Moreover, Pastry is resilient to concurrent node failures, and so is Squirrel. Upon failure of multiple nodes, Squirrel only has to re-fetch a small fraction of cached objects from the origin web server. Authors conducted the tests based on web trace data and showed that Squirrel is feasible, efficient, and comparable in performance to a dedicated web cache in terms of latency, external bandwidth and hit ratio. At the same time, Squirrel has the advantages of being inexpensive, highly scalable, resilient to node failures and requires little administration. Although there are some factors that can degrade the performance like the existence of popular content in a given area can put too much pressure on certain nodes and geographically distributed Squirrel networks with higher internal latency. The system was developed taking in consideration a web scenario in a well connected networks but can be a good start to developed a cache system in the scenario that is consider in this work.

Considering more dynamic scenarios the work of M.Legény et al. [16] is a good example, where authors used a cache system where nodes maintain the topology of their vicinity, in order to reduce the communication overhead in a fully distributed, dynamic self-organizing network and keep this data structure up-to-date throughout minor and rapid changes in the network (in previous work [15] authors had already shown that caching the overlay topology is an efficient method for reducing the communication overhead when the topology of the network is quasi-static). Their work focuses on biologically inspired, fully distributed self-organization algorithms for large overlay networks, with an emphasis on clustering and load balancing. For clustering purposes authors chose a family of algorithms called On-Demand Clustering (ODC), more specifically the Spyglass algorithm [15]. It has been shown that ODC results in an emergent self-organization behavior, i.e. clusters are formed and expanded when local demand for that rises. Authors used a model where a load balancing problem generates the demand for the clustering. Regarding the caching strategies authors consider: No caching; Reference algorithm; Static pre-caching; On demand cache. Authors tested the cache strategies (No Cache; On Demand Cache; On Demand Cache with Pre-Caching; Random Cache; Size Sensitive Random cache; Success Based cache; Change Based cache; Type Sensitive Change Based Cache) with a focus on the communication overhead and the clustering/load balancing performance. The results are not very conclusive because they showed that some cache strategies are more suitable than the others, depending of the evaluation criteria and the scenario. Despite that, all examined smart caching strategies beat the reference algorithms in terms of overall performance. Moreover experiments pointed out that fast topology change naturally stimulates self-organization, i.e. tends to bring matching, and not overloaded nodes into the direct neighborhood at no cost, resulting in possibly smaller clusters but better overall load-balancing and job processing curves. Despite the fact that this work is not been directly applied in opportunistic networks and it not take in consideration information centric concepts, it give good indications in terms of key areas: caching, self-organization and clustering.

In [19] authors proposed MIX, a cooperative caching system to improve data availability in mobile information systems. MIX scheme can be considered as an extension of COOP scheme, it resolves the problem of flooding that COOP schemes has met. Authors assumed that the cooperative caching architecture has only one server,also called the base station (BS), and many mobile users (MU). The communication between the BS and MUs is wireless link.

MU is the mobile device used to send requests to server and it can move freely in a cell or from a cell to another cell. In this model focus in cooperating data among MUs in network . Each MU has a service zone in which they can communicate with others. The connection among MUs is P2P wireless network. The conducted tests were focused on cache discovery cache admission and cache replacement and showed that MIX improved data availability and access performance. Cache hit ratio archived 70% and the average number of messages that MIX generates for a data request is much lower than COOP within 2 hops. Although authors did not prove the scalability of the model, they did not address the problem of cache consistency and considers the existence of good connectivity in the vicinity of the nodes (in disruptive scenarios the algorithm may not be adequate). The article enforces the importance of caching in mobile environment as an effective way of reduce latencies, and improve data availability.

2.2.3 General Conclusions

The existence of in-network storage for caching is one of the new and most important features present in all information centric architectures. Despite that, the best way of take advantage of this cache is still an open question. In the analyzed architectures and state of the art, the utilization of the cache is limited to store the transit information for a period of time, with cache replacement policies based on the age or usage of the content. Each cache work alone regardless of its neighbors (without any network awareness) in a passive way, i. e., do not cooperate actively to exchange information in order to optimize the local utilization of the caches and to enhance the data delivery. Caching techniques are not new and over the years they have been applied in various scenarios including, as shown, in mobile environments (ah-hoc networks). In these studies cooperative cache systems have been developed using self-organization concepts with good results. Although in these cases caching was implemented as an overlay systems and not as in-network cache in a information centric scenario, they present good indication about what can be achieved. Also self-organization is not a new concept, but its effectiveness is unquestionable in several scenarios including disruptive scenarios.

From the presented literature it's clear that there is no consensus among the information centric research community regarding the issues of naming, name resolution and routing. There is no definitive answer regarding if there will be a global solution to these problems or if the solution will be scenario dependent. Moreover there are some work done regarding information centricity and mobile scenarios (ad-hoc networks) but it is considered the existence of at least one end-to-end path among any pair of nodes at any time which may not be true on an opportunistic network and do not explore the full potential of the in-network caches. These facts highlight the need of find a suitable forwarding engine to opportunistic networks.

Self-organization and cooperative caching techniques will be explored in order to devise a cache system, in conjunction with the new forwarding engine to support it, that can improve data delivery and quality of service perceived by users in an information centric opportunistic network.

Chapter 3

Scientific Contributions

According to the state of the art and the mentioned goals of this work, to investigate the best information-centric networking approach for opportunistic networks in scenarios with heterogeneous devices, both in terms of network capabilities and system resources (CPU, memory, etc), the proposed contribution ***focus on the self-organization of in-network caches and a declarative forwarding system to exchange information among caches based on perceived information usage patterns***.

The major motivation for the established goal is the fact that the analysis of state-of-the-art shows a focus of prior-art on forwarding schemes aiming to transport information between producers and consumers and the naming scheme needed to find information. In almost all approaches information is only moved in the network when, and only when, an end consumer shows some interest on such piece of information. Moreover, few work has been done to analyze the best way to allow caches to cooperate and to allow information to flow among them in order to better serve future requests.

The investigation of self-organized systems of in-network caches is important because information-centric architectures embraced the fact that caching helps balance traffic, enhance the perceived quality of service and ultimately can help reduce usage and operational costs, and made it part of the network model (instead of be some kind of overlay). This is especially true in networks formed by mobile personal devices, where the connectivity can be intermittent: the bandwidth is limited and the resources are more limited.

To support an asynchronous exchange of information among caches, concepts of information naming and forwarding need to be investigated, namely the selection of the best naming and resolution strategy (e.g. late binding or

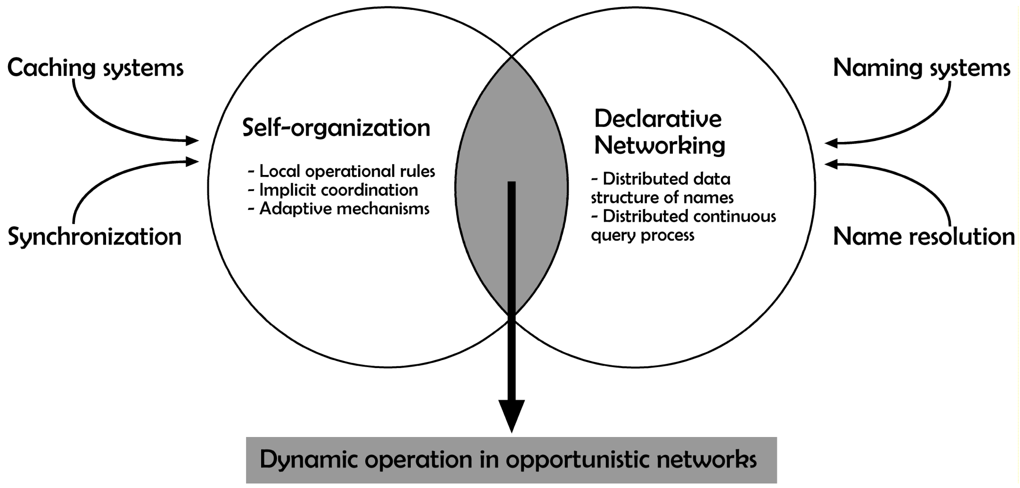

Figure 3.1: Major contributions

step-wise resolution of flat or hierarchical names) to support a declarative forwarding strategy among caches.

The fundamental paradigms that will drive this work are the self-organization and the declarative routing. The starting points, as described in the state of the art, are the existing caching systems, self-organized systems, naming and name resolution schemes. As shown in Figure 3.1 the ultimate goal is to achieve a dynamic operation in opportunistic networks.

Hence this thesis aims to contribute to find the answers to the following questions:

1. Aiming to devise a self-organized system of in-network caches:

   - How to use community detection techniques (clustering) to detect caches that have the same usage patterns in order to predict future demands?
   - How to synchronize caches in order to achieve an efficient scheduling of information?

2. Aiming to devise a declarative information networking system among caches:

   - What is the best naming and resolution strategy (e.g. late binding or step-wise resolution of flat or hierarchical names)?
   - How to use declarative forwarding strategy among caches in order to achieve an efficient exchange of information?

The remaining of this section provides more detailed insight about the motivation and goals of the two identified contributions of this thesis to devise efficient information-centric opportunistic networks.

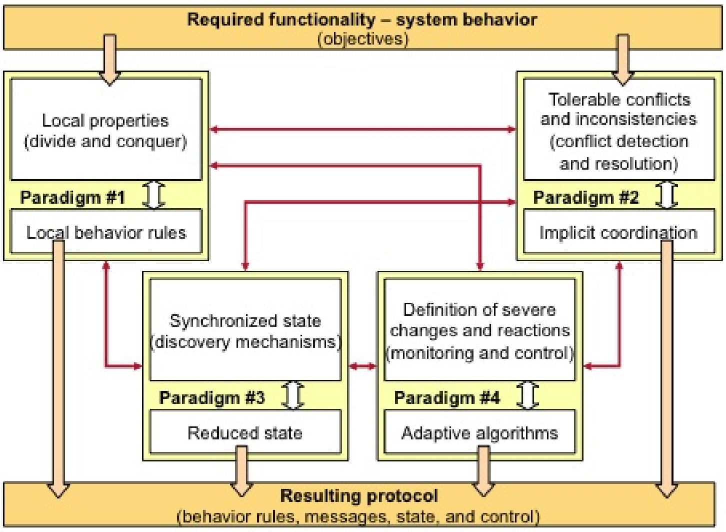

Figure 3.2: Self-organized in-network caching system

3.1 Self-organization of in-network Caching System

In an information-centric network the set of autonomous (deployed controlled by different entities) and heterogeneous in-network caches constitutes a complex system: a system of many parts which are coupled in a nonlinear fashion. While a linear system is subject to the principle of superposition, and hence is literally the sum of its parts, a nonlinear system is not. When there are many non linearities in a system (many components), its behavior can be very unpredictable. In this context, self-organization is presented as a solution for management and control of dynamic, highly scalable, and adaptive in-network caching systems.

Self-organization of in-network caches is a process in which structure and usage pattern at the global level of a in-network caching system emerge solely from numerous interactions among the lower-level caches without any external or centralized control. The system's components (in-network caches) interact in a local context by means of direct communication of environmental observations without reference to the global pattern.

The first contribution of this thesis is on the investigation of the properties of an self-organized systems of in-network caches, based on the four paradigms illustrated in Fig. 3.2, aiming to identify local operational rules for caching, methods for implicit coordination to detect conflicts and reduce state, and adaptive algorithms able to react to changes in cache usage patterns.

The devised self-organized system will operate based on a set of behavior rules and state, which coordination in the global information-centric network will be done based on declarative networking functions, described in the next sub-section.

3.2 Declarative Forwarding System

This thesis is focus on the vision of an emergence of a new type of communication based on physical proximity, where people encounter each other and devices directly communicate taking advantage of any communication opportunity within their range. Building an efficient information-centric dissemination and a search mechanism for such opportunistic networks are highly influenced by interaction of people, mobility, distribution of information in in-network caches.

On the other hand, after 30 years of Internet operation and research, running packet networks remains a complex and difficult problem, mainly because no abstractions have emerged to modularize the problem. If we look at a more abstracted level, the set of routing tables in a network represents a distributed data structure characterized by a set of ideal properties which define the network. If the thinking is in terms of structure, not protocols, routing is the process of maintaining these properties in the face of changing ground facts (e.g. failures, topology changes, load, policy). Looking at the case of information-centric networking and the usage of in-network caches, routing tables are a view over changing information distribution conditions and state. Maintaining such routing structure is the domain of distributed continuous query processing.

Hence, one goal of this thesis is to contribute to define a constrained declarative language for opportunistic networks [26, 14, 8], to allow a higher-level view of information-centric routing properties. This means that routing is not seen anymore as a simple configuration language, allowing a modular decomposition of functions, such as name resolution, and a dynamic optimization of processes, such as the support for self-organized in-network caching systems. This optimization is possible since, a declarative structure off routing may make it easier to incorporate other knowledge into routing policies, such as the the natural integration of discovery functions.

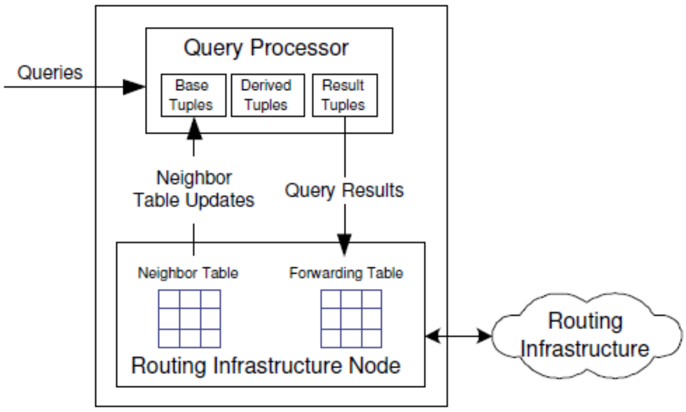

Figure 3.3: Basic Components of declarative routing Infrastructure [14]

Chapter 4

Experimental Methods and Tools

Throughout this thesis, the experimental work will be based on a identified benchmark architecture, which should be the one presenting the better characteristics to embedded the contributions of this thesis.

In a first phase it is expected the evaluation of different solutions for the proposed self-organized system using Matlab. In a second phase, the evaluation of declarative forwarding algorithms may be done also based on Matlab (to focus on the algorithm) or based on a implementation on a large testing system such as Planetlab, which will allow the emulation of a real system in which some Internet nodes will be able to run information-centric networking processes.

After obtaining enough evidence from simulation, a prototype of the devised solution should be designed in order to test it in a real deployment to have a better understanding of its performance. Although this is not a mandatory goal, the demonstration of the proposed systems is expected to be done based on a ***cooperative sensing scenario*** where users carry personal devices with communication interfaces (e.g. 3G, Wi-Fi and Bluetooth) and the capability of sensing the world through internal sensors. These devices will be able to extract real-world information that will be disseminated based on a information-centric platform encompassing the contributions of this thesis. In such scenario, users are highly mobile and can move into areas with poor connectivity. The assumption is that the personal device has a very acceptable CPU power and good storage capabilities and will be able to communicate with any devices that appears in its vicinity. The devised scenario may be complemented with the existence of embedded devices spread out to several locations that users can interact when passing in their vicinity. These devices have limited CPU power and limited storage capability but have a great potential to incorporate all kind of sensors. In the devised scenario, the above mentioned devices will cooperate, in a opportunistic way, to achieve a large scale sensing experience. The cooperation will be performed in self-organized way, without the existence of any central coordination.

Chapter 5

Work Plan and Methodology

This document describes the result of the pre-proposal phase, which aimed to prepare the candidate towards a set of important research topics as well as towards the thesis roadmap needed to achieve international accepted results. This pre-proposal phase was based on brainstorming sessions aiming to identify the right motivation and problem statement. The pre-proposal phase was also marked by discussions about research techniques, evaluation tools and the most important scientific conferences, magazines, and journals in the networking field.

The work developed in this PhD is divided into two stages: the pre-proposal stage, and the pos-proposal stage. The pre-proposal stage aimed to prepare the thesis plan and roadmap, ensuring optimal goals throughout the PhD course, including a good analysis of related solutions. Such work is to be followed, in the pos-proposal stage, by two stages corresponding to the investigation of the proposed contributions: self-organized in-network caching system and declarative forwarding system. Each one of these two technical phases will use the same work method: start with a design phase aiming to delve into potential improvements, both from a conceptual perspective as well as from an analytical perspective. A specification phase aims to detail novel algorithms, and/or protocols, which will be evaluated by recurring to Matlab and Planetlab whenever necessary. The two technological stages will be followed by a prototyping stage, in which the evaluated solutions will be incorporated in a local test-bed of a pre-identified information-centric architecture and demonstrated, and if possible demonstrated based on a cooperative sensing scenario.

- Phase 1: **Brainstorming and pre-validation** (March 1st, 2011 - September 15th, 2011)
  - Expected outcome : Analysis of related work aiming to identify the generic properties and challenges of current information-centric architectures, as well as the pos and cons of major proposals in the area of: i)

self-organized systems; ii) routing and name resolution. Further detail the appointed challenges and to validate solutions found in the literature, ensuring optimal goals throughout the PhD course, including a good analysis of related solutions. This stage is already in progress with the detailed comparison of the CCN and HAGGLE architectures, which have been identified as good representatives of an alternative Internet architecture (CCN) and an information-centric approach for more disruptive scenarios (HAGGLE);

  - Expected publications: 1 PhD thesis proposal, 1 workshop paper (e.g. IEEE ICC);

- Phase 2: **Development of a self-organized in-network caching system** (October 1st, 2011 - June 30th, 2012)
  - Expected Outcome: Investigation of self-organized models in order to evaluate how suitable are they for operation over opportunistic networks and to support the devised operation of in-network caches. Such initial finding will be used to derive algorithms to: identify local operational rules for caching based on know interested for information; methods for implicit coordination in order to detect conflicts and reduce state (e.g. synchronization of information instances); and adaptive algorithms able to react to changes in cache usage patterns;
  - Expected publications: 1 technical report; 2 conference publications (e.g. ACM Mobicom, ACM CoNext);

- Phase 3: **Development of a declarative forwarding system** (July 1st , 2012 – March 30th, 2013)
  - Expected outcome: Investigation of a more modular approach routing and name resolution aiming to provide the abstraction needed to modularize the communication problem over complex systems as the ones created out of a large number of in-network caches. This investigation encompasses: i) definition of a distributed structure of references to named-data (flat or hierarchical based) characterized by a set of ideal properties which define the information network; ii) distributed continuous query processing (including name resolution process) to face changing ground facts (e.g. failures, topology changes, information load, information interest policy);
  - Expected publications: 1 technical reports; 2 conference publications (e.g. ACM MobiHoc, IEEE Infocom);

- Phase 4: **Overall Evaluation** and Demonstration (April 1st, 2013 - October 31th, 2013)
  - Expected outcome: Combined evaluation of the self-organized and declarative forwarding systems, based on simulators and PlanetLab and possibly in a local implementation of a cooperative sensing scenario;
  - Expected publications: 1 journal paper (e.g., IEEE/ACM TON, IEEE JSAC); 1 thesis.

To achieve the expected levels of success (cf. Chapter 6) the work plan will be coordinated based on weekly meetings between the advisor and the student to discuss work progress and next steps. The candidate is expected to deliver several intermediate reports to show the progression of the technical work. Each report will be provided in English. Moreover, it is expected the candidate to give periodically presentations in his host institution as well as in other locations and events, in order to collect comments from relevant researchers. The final outcome of the thesis work will be provided in the form of a technical report and a presentation due in 2013/2014. Assumptions, design decisions, configuration choices, and results are part of the report as well as usage information. Correct bibliographic references and a list of papers, recommendations, and descriptions used must be added, including the ones listed in this description.

The first phase ("**Brainstorming and pre-validation**") consisted mainly of the analysis of state-of-the-art solutions to gather enough knowledge about information-centric architectures and self-organized systems. It is worth pointing out that this state-of-the-art analysis can last up to the design phase as new proposals are expected to emerge during the development of this thesis. From the first phase forward, the student is expected to have enough knowledge to present to and discuss with the advisor in regular meetings. This regular presentations and meetings will help keeping the student on the correct track throughout the development of the thesis work.

It is expected that during the second and third phases the candidate will periodically revisit assumptions and requirements, which may lead to a new analysis of related work and solution finding, followed by the needed update of the specification and validation. Nevertheless, the main roadmap should be kept as described above.

Next, the final chapter summarizes the scientific milestones expected to be achieved with the the development of this thesis work.

Chapter 6

Scientific Milestones

Throughout the thesis, the following (minimum) success indicators are expected:

- 2 technical reports (encompassing all the finding in relation to self-organization of caches, as well as declarative forwarding methods;
- 1 workshop publication (related to the first phase)
- 4 conference publications (related to the second and third phase)
- 1 journal paper (related to the four phase, and aiming to archive all the major findings)

At least two conference papers and the journal paper are expected to be accepted in top-level events from IEEE or ACM, such as ACM CHANTS, IEEE INFOCOM, IEEE ICC, ACM SIGCOMM, ACM MobiHoc, IEEE Journal on Selected Areas in Communications, and IEEE/ACM Transactions on Networking, as indicated in the workplan. Optional venues for the publication of scientific results are IEEE Globecom, IEEE ICNP, ACM HotNETs, Elsevier Communication Magazine, ACM Computer Communications Review and IEEE Communication Magazine. The final outcome of the thesis work will be provided in the form of a technical report, a presentation and a set of software modules to be made available to the research community.

Besides the publication of scientific work in major international venues, it is expected the participation in several national events that aim to disseminate research work, such as the Portuguese support action Approaches to Paradigms of a future Internet (API), and the realization of demonstrations in national events.

Bibliography

[1] Andrew Berns and Sukumar Ghosh. Dissecting self-* properties. In *Proceedings of the Third IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, pages 10–19, Washington, DC, USA, 2009. IEEE Computer Society.

[2] S. Bhattacharjee, K.L. Calvert, and E.W. Zegura. Self-organizing wide-area network caches. In *INFOCOM. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 600 – 608, mar-2 apr 1998.

[3] Robert Braden, Ted Faber, and Mark Handley. From protocol stack to protocol heap: role-based architecture. *SIGCOMM Comput. Commun. Rev.*, 33:17–22, January 2003.

[4] Amos Brocco, Apostolos Malatras, and Béat Hirsbrunner. Enabling efficient information discovery in a self-structured grid. *Future Gener. Comput. Syst.*, 26:838–846, June 2010.

[5] Kevin Cheng, Ali Akbar Zarezadeh, Felix Muhlbauer, Camel Tanougast, and Christophe Bobda. Auto-reconfiguration on self-organized intelligent platform. In *Adaptive Hardware and Systems, NASA/ESA Conference on*, pages 309 –316, 2010.

[6] C. Dannewitz and T. Biermann. Prototyping a network of information. In *34th IEEE Conference on Local Computer Networks*, pages 1 –6, October 2009.

[7] C. Dannewitz, J. Golic, B. Ohlman, and B. Ahlgren. Secure naming for a network of information. In *INFOCOM IEEE Conference on Computer Communications Workshops*, pages 1 –6, march 2010.

[8] Timothy G. Griffin and João Luís Sobrinho. Metarouting. In *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communications*, pages 1–12, New York, NY, USA, 2005. ACM.

[9] Sitaram Iyer, Antony Rowstron, and Peter Druschel. Squirrel: a decentralized peer-to-peer web cache. In *Proceedings of the twenty-first annual symposium on Principles of distributed computing*, pages 213–222, New York, NY, USA, 2002. ACM.

[10] Van Jacobson, Diana K. Smetters, Nicholas H. Briggs, Michael F. Plass, Paul Stewart, James D. Thornton, and Rebecca L. Braynard. Voccn: voice-over content-centric networks. In *Proceedings of the workshop on Re-architecting the internet*, pages 1–6, New York, NY, USA, 2009. ACM.

[11] Van Jacobson, Diana K. Smetters, James D. Thornton, Michael F. Plass, Nicholas H. Briggs, and Rebecca L. Braynard. Networking named content. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pages 1–12, New York, NY, USA, 2009. ACM.

[12] Teemu Koponen, Mohit Chawla, Byung-Gon Chun, Andrey Ermolinskiy, Kye Hyun Kim, Scott Shenker, and Ion Stoica. A data-oriented (and beyond) network architecture. In *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communications*, pages 181–192, New York, NY, USA, 2007. ACM.

[13] Zhenping Li, Ruisheng Wang, Xiang-Sun Zhang, and Luonan Chen. Self-organizing map of complex networks for community detection. *Journal of Systems Science and Complexity*, 23:931–941, 2010. 10.1007/s11424-010-0202-3.

[14] Boon Thau Loo, Joseph M. Hellerstein, Ion Stoica, and Raghu Ramakrishnan. Declarative routing: extensible routing with declarative queries. In *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communications*, pages 289–300, New York, NY, USA, 2005. ACM.

[15] B. K. Benk M. Legeny. Design of novel self-organization algorithms through simulation. In *Proceedings of EUROSIS*, pages 10–15, 2010.

[16] B. K. Benko M. Legeny. Topology cache aided self-organization for ad-hoc mobile networks. *Cyber Journals: Multidisciplinary Journals in Science and Technology - Journal of Selected Areas in Telecommunications (JSAT)*, pages 56–63, March 2011.

[17] Michael Meisel, Vasileios Pappas, and Lixia Zhang. Ad hoc networking via named data. In *Proceedings of the fifth ACM international workshop on Mobility in the evolving internet architecture*, pages 3–8, New York, NY, USA, 2010. ACM.

[18] Lixia Zhang Michael Meisel, Vasileios Pappas. Listen first, broadcast later: Topology-agnostic forwarding under high dynamics. In *Annual Conference of International Technology Alliance in Network and Information Science*, page 8, September 2010.

[19] Thu Nguyen Tran Minh and Thuy Dong Thi Bich. An efficient model for cooperative caching in mobile information systems. In *Proceedings of the IEEE Workshops of International Conference on Advanced Information Networking and Applications*, pages 90–95, Washington, DC, USA, 2011. IEEE Computer Society.

[20] Renato E. Mirollo, Steven, and H. Strogatz. Synchronization of pulse-coupled biological oscillators. *SIAM J. Appl. Math*, 50:1645–1662, 1990.

[21] Erik Nordström, Per Gunningberg, and Christian Rohner. A search-based network architecture for mobile devices. Technical Report 2009-003, Uppsala University, Computer Systems, 2009.

[22] Soon Y. Oh, Davide Lau, and Mario Gerla. Content centric networking in tactical and emergency manets. In *Wireless Days, IFIP*, pages 1 –5, October 2010.

[23] Antony I. T. Rowstron and Peter Druschel. Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems. In *Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms Heidelberg*, pages 329–350, London, UK, 2001. Springer-Verlag.

[24] Jing Su, James Scott, Pan Hui, Jon Crowcroft, Eyal De Lara, Christophe Diot, Ashvin Goel, Meng How Lim, and Eben Upton. Haggle: seamless networking for mobile applications. In *Proceedings of the 9th international conference on Ubiquitous computing*, pages 391–408, Berlin, Heidelberg, 2007. Springer-Verlag.

[25] A. Tyrrell, G. Auer, and C. Bettstetter. Emergent slot synchronization in wireless networks. *Mobile Computing, IEEE Transactions on*, 9(5):719 –732, May 2010.

[26] Eiko Yoneki, Ioannis Baltopoulos, and Jon Crowcroft. D3n: programming distributed computationin pocket switched networks. In *Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds*, pages 43–48, New York, NY, USA, 2009. ACM.