1. MOTIVATION.

Distributed Systems have changed substantially in the recent past with the digital revolution epitomized by the phenomena of social networks and cloud computing. In the previous incarnation of distributed computing [16] the emphasis was on consistency, fault tolerance, resource management and related topics; these were all characterized by interaction between processes. Research proceeded along two lines: the algorithmic side and the more process algebraic approach where the emphasis was on developing compositional reasoning principles. What marks this new era of distributed systems is an emphasis on managing access to information to a much greater extent and flexibility than before.

Problem. One of the most fundamental concerns in today’s distributed systems is that of privacy. Particularly in social networks and cloud storage where agents (users) continuously post beliefs/facts (e.g., opinions, activities, geographic position) and execute processes (applications) accessing their local spaces (e.g. biography, address book and albums). This may give malicious agents the possibility of gathering private information often without the user even being aware of it. Moreover, privacy policies set by a user may become unintentionally violated by policy changes of the social network administration without the user or administrator awareness. Another central concern, closely related to privacy, is that of reliability. A user may need guarantees that a process behave as intended, particularly if it is executed in his own local space, or may need to be sure that processes running externally does not interfere with processes in his local space. Defective or malicious programs may disrupt computer operation or gain access to sensitive information.

New Aspects. We now wish to underline two prominent aspects of this new era of distributed systems that are central to this proposal. The first one is the intrinsic epistemic nature of these systems arising from the presence of social behaviour. We constantly have millions of users posting facts, opinions, beliefs and making decisions on social networks; (epistemic) information is everywhere. How precisely and by whom a user's opinions and beliefs can be seen may have a significant impact on his social/professional life. Being able to model this epistemic flow in a highly distributed environment would thus be an important contribution to assess privacy breaches risks. The other is the spatial nature of computation. With the advent of cloud
computing and other digital services, computational space is now more than ever available to everyone; space is everywhere. Users, programs and information are spatially distributed in these networks. A solid understanding of this notion of space is fundamental in any model of today’s distributed systems.

**Epistemic reasoning.** Epistemic concepts were crucial in distributed computing as was realized in the mid 1980s with Halpern and Moses’ groundbreaking paper on common knowledge [13]. This led to a flurry of activity in the following few years [11] with many distributed protocols being understood from an epistemic point of view. However, the impact of epistemic ideas in the concurrency theory community has been slower in coming. Social networks is a new challenging phenomenon of great interest to concurrency theorists where these ideas can be exploited due to their inherently epistemic nature: multiple agents interacting by posting and querying information in terms of beliefs and knowledge. Epistemic concepts such as public announcements, shared and distributed information have natural corresponding phenomena in these networks.

**Spatial reasoning.** There is a large body of methods and techniques to deal with space in logic and several fields of computer science (e.g., including non-monotonic and modal logics, circumscription methods, chronological minimization methods, relation algebras, and applications of constraint-based reasoning.). However, despite significant progress by the concurrency theory community [4,5,6], the use of formalism and techniques to describe the spatial structure of distributed and mobile systems is still in its infancy. Today’s distributed systems, distributed databases, mobile systems, geographic information systems, tracking systems and other complex systems from the digital revolution challenge the capabilities of existing techniques and representation methods for spatial reasoning. Familiar services such as shared albums, data and applications in social networks and cloud storage, provide natural examples of managing information access in the presence of spatial hierarchies. These domains pose important problems such as the design of models to predict and prevent privacy breaches, which are commonplace currently.

2. **GOAL.**

Given their complexity and increasing popularity, the design, analysis, simulation, and programming of the new era of distributed systems will continue to raise important challenges to computer science.

We propose to rise to this challenge by developing an innovative and expressive computational model for these systems that will coherently combine techniques for the analysis of concurrent systems such as process calculi with epistemic and spatial formalisms. The model should provide reasoning techniques to predict potential privacy breaches as well as intrusive and unreliable behaviour.

Epistemic analysis will be used to reason about the distributed agents’ information, beliefs and
knowledge acquired during computation. Spatial formalisms will be used to describe distributed network topologies. The integration of these formalisms in a single framework will allow us to capture meaningful families of modern distributed systems that cannot be faithfully modeled in any previous concurrent framework.

3. APPROACH.

*Hierarchical Distribution of Space.* To capture significant families of today’s distributed system we shall focus on multiple agents sharing information in the presence of hierarchical distribution of the space. In our setting a *spatial multi-agent system* consists of a space (a *cloud*) that may contain data (information in terms of facts or beliefs), processes (*applications*), and other spaces (its *sub-spaces*) each of which is itself a system. Nesting of spaces may be of any depth. Each space within the system is attributed to an agent. In its space, the agent may store information, run applications, and attribute sub-spaces to other agents. This, for example, captures the familiar situation where a user can grant computational spaces on his computer to other users.

Figure 1 illustrates a simple example of a three-agent system with hierarchical distribution of the space describing some typical scenarios. The owner of the outermost cloud is omitted. Agent 1 owns a space (the leftmost red cloud) in which it grants a space to Agent 3. Agent 2 and 3 share some applications and data.

*The D-Spaces Model.* Process calculi are linguistic formalisms from concurrency theory for expressing and analyzing concurrent systems. Nevertheless, the systems we identify have
aspects that cannot be expressed with the traditional process calculi, and require the introduction of new concepts. In particular, we need to express agents posting and querying information in presence of spatial hierarchies for sharing information and knowledge.

In [26], the advisor and co-advisor of this proposal, along others, put forth a process calculus, $sccp$, featuring two novel operators: a spatial and an epistemic operator. The spatial operator may specify a process, or a local store of information, that resides within the space of a given agent (e.g., an application in some user’s account, some private data shared with a specific group). The epistemic operator may specify that the information computed by a given process will be known to a given agent. To the best of our knowledge, other process calculi can only express these epistemic concepts and the spatial distribution of information indirectly.

**Research Strategy.** Although our $sccp$ model represents a significant first approach, in its present form it is not sufficiently robust for the today’s distributed systems as we argue below. Our research strategy is to develop a conservative extension of $sccp$ that captures more faithfully the systems under consideration. We now identify three crucial and technically challenging phenomena, not included in $sccp$, that we shall explore in the development of our model. Henceforth we refer to the proposed extension as $D$-$Spaces$.

1. **Spatial mobility.** The model $sccp$ does not capture mobility, a common feature in the systems under consideration: Agents and programs may change from one space to another. Mobile systems now pervade the digital world. Their distinguishing property is that their components, in our case agents and applications, can change their communication structure, move around and exchange information. Other examples include GPS applications running on mobile devices, and in general agents in distributed systems with dynamic and hierarchical link topologies. We plan to adapt the notion of “change” from logic frameworks to capture mobility in $D$-$Spaces$. By using ideas from temporal and dynamic logic [28,29] possibly in combination with the notion of mobility from process calculi such as Ambients [6] we expect to be able to express spatial change during the temporal evolution of distributed systems.

2. **Quantitative Reasoning.** $sccp$ does not include quantitative spatial and epistemic information. We propose an extension that will allow us to add weights to agents beliefs and information within the spaces of the system. Our approach is to generalize the $sccp$ notion of space in $D$-$Spaces$ as a form of probabilistic measure on epistemic statements to express meaningful and wider social scenarios. For instance, suppose that $c$ represents a statement, the perception of which may change according to space, time or the individual — e.g., an inherently subjective statement about politics, religion, or sports. We may have, for example, that agent $i$ initially believes that $c$ holds with some probability $p<1$. After interacting with other agents or moving across different spaces, agent $i$ may end up convinced that $c$ is true (or that it is false). We can also express that agent $i$ believes $c$ more than agent $j$ does and other similar statements related to social interaction. This quantitative modeling provides a more faithful representation of virtual
social behaviour in distributed networks, where information certainty changes parallel to its dissemination.

3. **Scalability.** Furthermore, sccp assumes a priori-fixed number of agents. Although this is a typical restriction in epistemic formalisms [11], it limits considerably our model. The number of users in popular social networks, cloud computing systems grows continually, hence it is virtually unbounded. This presents us with a technical challenge since our present results depend on the assumption of a fixed number of agents. Our approach is to build upon the ideas from nominal logic and the pi-calculus [19] whose intended phenomena is the creation of new links. We foresee the introduction of an operator for creating dynamically new agents.

4. **EXPECTED RESULTS.**

The expected outcome of this project is two-fold: Theoretical and practical results. On the theoretical side we will advance the state of the art of concurrency theory with a formalism that deals with new challenging concepts from epistemic and spatially distributed systems. In particular, mobile agents and programs, quantitative spatial information, and unboundedly many agents will be allowed in the modeled. The model will be also equipped with reasoning techniques to predict and prevent, by construction, privacy breaches and unreliable behaviour.

On the other, we expect to produce an automated tool to simulate and verify potential privacy breaches and unreliable behaviour on the systems under consideration. We shall also implement a prototype of a social network based on specialization of $D$-Spaces.

5. **BACKGROUND AND RELATED WORK.**

Our project relies on notions and tools from logic and concurrency theory for concurrent systems: process calculi, ambient and modal logics.

*Process calculi* provide a language in which the structure of terms represents the structure of processes together with a transition relation to represent computational steps. One of the main representatives of these calculi is the *pi-calculus* [19] whose intended model is concurrent systems with link mobility.

*Modal logics*, such as spatial epistemic logics [4,5,11], are used for specification of spatial and knowledge properties in mobile systems. For instance, it may be specified that there is an access route between two different sites, that an agent will cross the firewall, that there is a virus within the system, or that certain agents shared certain private knowledge. To do this, these logics include modalities to make assertions about system locations and their names as well as to make assertions about distributed and common knowledge. Spatial epistemic logics
have been used for the verification of spatial properties of mobile systems and security protocols [6,27].

There is a huge volume of work on epistemic logic and its applications to distributed systems; [11] gives a good summary of the subject. This work is all aimed at analyzing distributed protocols using epistemic logic as a reasoning tool but not as a part of the computational formalisms as in our proposal. Epistemic logic for process calculi has been discussed in [7, 9, 14]. In all of the above-mentioned works, however, the epistemic logic is defined outside of the process calculus, with the processes as models for the logic, whereas our processes have epistemic (or spatial) logic terms within the constraint system, as well as knowledge or space constructions on the processes.

The Concurrent Constraint Programming (ccp) process calculus [Saraswat, 1990] was designed for reasoning, modeling and programming concurrent and reactive agents that interact with each other and their environment by telling and asking information represented as logic formulae (constraints). This formalism enables a unified representation of processes (behavioral style) and temporal formulae asserting properties of the system evolution (logical style). In addition, it has been shown to be a flexible and versatile framework where several variants can be experimented with and validated. The ccp paradigm has been successfully used for applications with advanced interaction with their environment and given rise to a very active research field. The sccp formalism [26], already discussed in the Approach, extends ccp with the ability to express epistemic concepts.

The issue of extending ccp to provide for distributed information has been previously addressed in [23]. In [23] processes can send constraints using communication channels much like in the pi-calculus. This induces a distribution of information among the processes in the system. This extension, however, is not conservative with respect to ccp and hence does not share the goal of the present proposal. Another closely related work is the Ambient calculus [6], an important calculus for spatial mobility. Ambient allows the specification of processes that can move in and out within their spatial hierarchy. It does not, however, address posting and querying epistemic information within a spatial distribution of processes. Adding ‘Ambient’-like mobility to D-Spaces is a natural research direction.

The authors of [30,31] use probabilistic methods for ccp calculi. The intended phenomena they wish to capture is random choice and assignment. The phenomena we would like to model instead is social behaviour, in particular agent’s beliefs. Therefore, there is no clear connection between the goal of the present proposal and the work in [30,31].

One very interesting approach related to ours in spirit – but not in conception or details – is the spatial logic of Caires and Cardelli [4, 5]. In this work they also take spatial location as the fundamental concept and develop modalities that reflect locativity. Rather than using modal logic, they use the name quantifier which has been actively studied in the theory of freshness of names in programming languages. Their language is better adapted to the calculi for mobility
where names play a fundamental role. In effect, the concept of freshness of a name is exploited to control the flow of information. It would be interesting to see what a name quantified D-Space calculus would look like and to study the relationship with the Caires-Cardelli framework.

The process calculi in [2, 3, 10] provide for the use of assertions within pi-like processes. They are not concerned with spatial distribution of information and knowledge. These frameworks are very generic and offer several reasoning techniques. Therefore, it would be interesting to see how the ideas developed in the proposal can be adapted to them.

The seminal work in [32] presents a framework for analyzing privacy and anonymity in social networks and a new re-identification algorithm targeting anonymized social-network graphs. The approach in [32] is based on graph-theory and probabilistic methods. Our proposal is also concerned with privacy but it is based instead on concurrency, epistemic and spatial formalisms. As such we are interested in the behavioral understanding of spatially distributed systems, an issue not addressed in [32].

REFERENCES.


**Responsible:**

Dr. Frank Valencia - LIX, Ecole Polytechnique.