UCD Systems Research Group

Like reality, but different…

Autonomic networking: Achieving stability in the face of pervasive uncertainty

Simon Dobson

Systems Research Group
School of Computer Science and Informatics
UCD Dublin, Belfield, Dublin 4, Ireland

http://www.ucd.ie/csi

simon.dobson@ucd.ie

http://www.simondobson.org

Invited talk at Autonomic Networking 2006, Paris FR, 29 September 2006

Introduction

Autonomic communications is based on a tension

- Apply adaptive techniques to optimise the networks' and applications' behaviours in a changing situation
- Too much visible adaptation causes confusion, so we need to maintain a level of predictability and simplicity at the users' level

Aside from any individual technique or technology, we therefore have a large-scale systems problem

- Balancing chaos against order?

My goals in this talk

- To argue (again) for a semantics-driven approach to autonomics
- To present some ideas about continuity and compositionality in autonomic systems design

Overview

1. Autonomic and traditional networking
   - …and why we're here
2. The semantics of networked applications
   - …and how little we may actually know
3. The space of possible solutions
   - …and how they will always all be wrong
4. More structured approaches
   - …and how we might gain a measure of predictability
5. A call to arms
   - …and a plea for a more universal and well-founded ^ approach

…and well-funded

…and three controversies and five jokes

Traditional networking

We're engaged in a battle with the complexity of the systems that we're seeking to build

- …and we're losing…
- While the end-user experience has become easier – given the features on offer, anyway – there hasn't been a corresponding simplification in development, deployment or management
- These complexities affect the economics of new products, making some services non-viable

A good example might be VOIP hand-off across multiple WiFi networks with different management regimes

Consequences

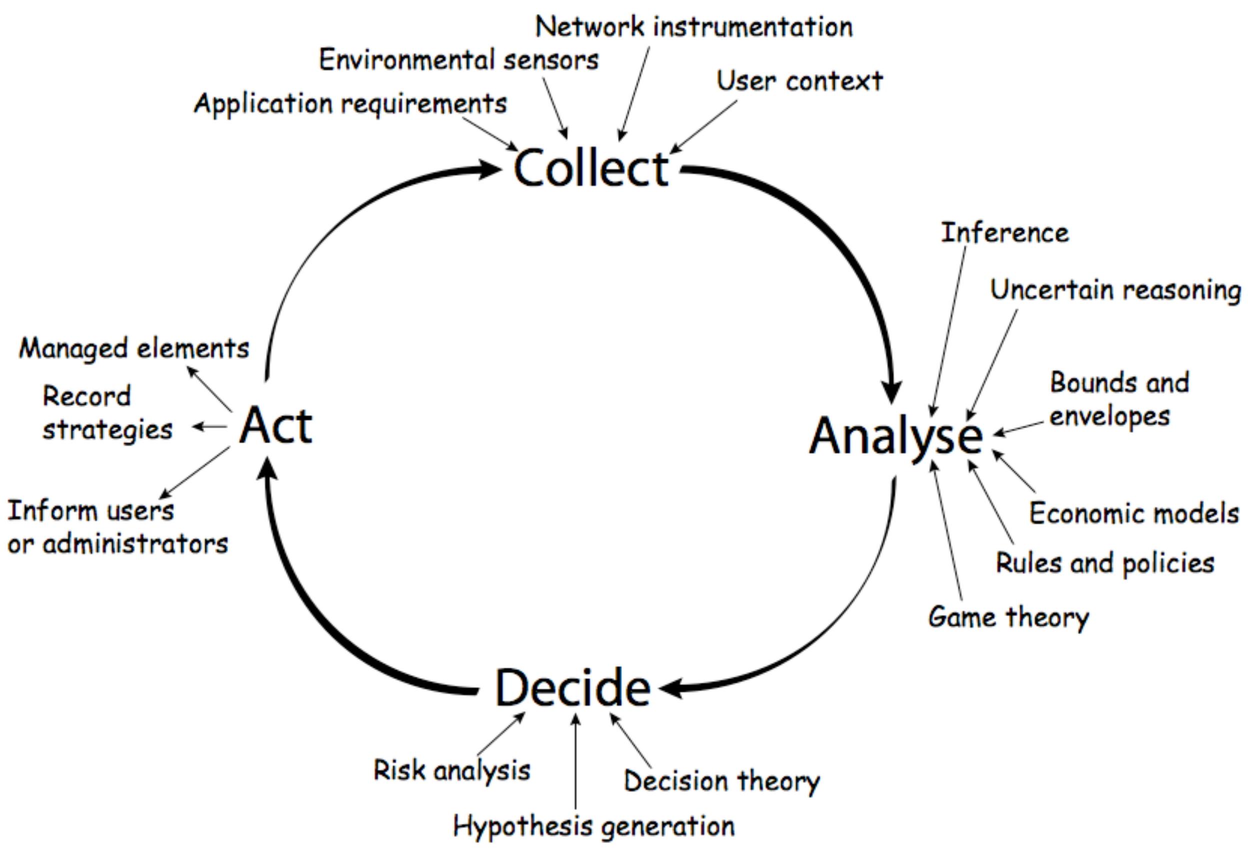
- We're being asked to make technical choices that we simply *can't make accurately*
- Decisions may fossilise parts of the network so that future changes are uneconomic or technically infeasible

Stability in the face of falling objects

In a sense, autonomic communications is about reducing the variability that passes from the environment to the application

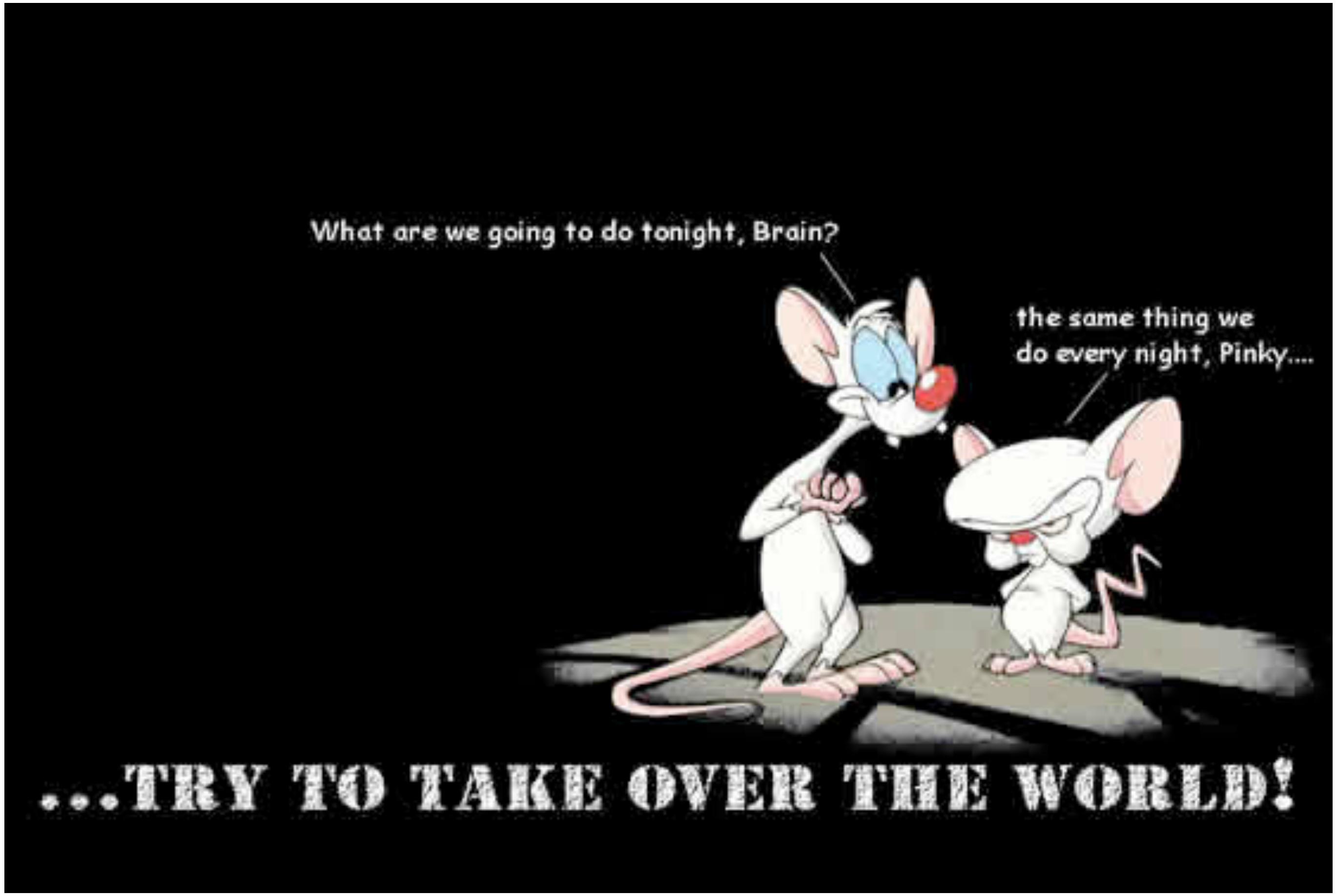
- Simplify management – self-configuration, self-healing
- Smooth behaviour – self-adaptation, self-optimisation
- Leverage and integrate information at different semantic levels

Autonomics



From Dobson, Denazis, Fernandez, Gaiti, Gelenbe, Massacci, Nixon, Saffre, Schmidt and Zambonelli. A survey of autonomic communications. ACM Trans. Autonomous and Adaptive Systems **1**(2). 2006. To appear.

Our place in the world…

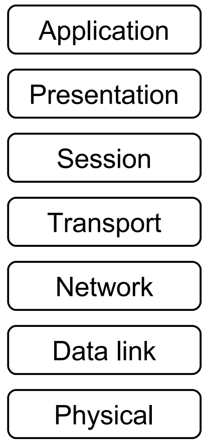


Pinky and the Brain. Copyright © Steven Spielberg and Warner Brothers

The difference

If we want these properties, we have to modify some of our underlying assumptions

- Classical (Shannon) information-theoretic view of the network as an uninterpreted bit stream – we want to leverage *meaning*, to handle different flows differently
- The clear separations of concerns across the layers can impede the free flow of metadata



"People design networks to send packets accurately"

But this implies integrating information with wildly differing characteristics

- Sensing digital *versus* sensing physical *versus* predicting people are all very different in terms of their accuracy and reliability
- This lies at the core of the autonomics problem

What do we mean by “meaning”?

What is the “meaning” of a movie?

- It’s a sequence of bits encoding a sequence of frames, whose encoding may “prefer” certain handling
- It’s to be watched, and so needs to be delivered in a way that allows it to be played with no delays or dropped frames
- Watching has a priority relative to other activities, like VOIP and fire alarms
- It’s to be stored, in which case it’s just a large block of bits
- It’s storage and transmission are governed by DRM restrictions

The “correct” adaptation potentially involves all these aspects of communications, on an on-going basis

Uncertainty

At this point we have to ask whether we can collect the information we need to make the decisions

- Local network load? Remote network load?
- The use the connection is serving?
- Other concurrent usage?
- The expected future evolution of these aspects?

So many of our decisions are going to be made on the basis of information that is *intrinsically* uncertain

- Improved instrumentation won't always help
- Location and other physical sensors are inherently noisy
- Human behaviour only follows a statistical distribution ahead of time

Context: the environment in which an activity occurs, understood symbolically

Making models – 1

We talk about self-* properties

- Self-management, self-optimisation, self-healing, …

What does it mean for a network to self-*anything*?

- Consider self-optimisation as an example
- You optimise *against* something: there are some criteria against which you seek to demonstrate optimal behaviour
- This (and other) self-* behaviours are generally defined relative to some external and human-focused model

So we arrive at a position in which we are solving autonomic problems by building a model of the world and the tasks within it, testing hypotheses against that model, and using this to guide decisions

Making models – 2

Of course, model-making itself is quite a complex task

- Compositional – describe different aspects separately
- Entwined – addressing one issue may involve several aspects of a model, and may have a knock-on effect on other issues
- Concurrent – a communications system is almost always engaged in several interactions simultaneously
- May be implicit – not reified explicitly

Is it going too far to suggest that this semantic multiplicity is the essence of autonomics?

- Can we claim that, in addressing a single issue, we are providing something autonomic? Is one level of meaning enough?

My heart rate does not depend solely on my exercise level, and directly influences my rate of breathing and so on

Stability

Let's consider a simpler problem: maintaining a video bit stream while mobile

- So we want to move prioritise packets such that we have a sufficiently buffered stream to maintain uninterrupted playing
- Well-addressed both in theory and experimental practice

Actually a surprisingly subtle problem

- Very sensitive to changes in network characteristics, and to the activities of other users
- Affected by user habits and desires
- Renders the properties of network traffic perhaps uniquely visible

What sorts of "autonomic" strategies are available to us?

A space of solutions

There are a whole set of transmission regimes that would address this problem adequately

- Download the whole movie ahead of time and play locally
- Burst-transmit large blocks of data and cache locally
- Stream data packets only slightly ahead of use

Which one is "correct"?

Leverage high bandwidth and minimise local storage → …*if* we have such bandwidth available

Use the network's peak bandwidth effectively, keep the channel clear over long periods → …*if* we can grab it sufficiently often

Makes good use of forward knowledge and (probably) fixed-wire infrastructure → …*if* this knowledge is available

Bad idea 1: ahead of time

We could choose one of these solutions ahead of time

- Typically stream with variable buffering at the client

But this may be a remarkably inefficient choice

- User may have predictable habits or be engaged in a well-defined workflow in which video pre-loading is possible
- …or may be in a naturally "bursty" environment that would lend itself perfectly to a different strategy

The point isn't to argue about which *one* we pick, but rather to argue that *any one* can be sub-optimal, and can *become* sub-optimal over time

- Ideally we'd like to cover the design space
- Take information from wherever is comes to hand

Bad idea 2: defer to applications

We could instead choose to leave the decision to the application, and have it decide on the caching or streaming strategy

- How do we know they'll do it right?
- Will a local choice be non-locally (let alone globally) optimal?
- What are the implications for other users?
- May be hard for a small provider to get the necessary administrative permissions for widely-distributed applications

An application probably has no better chance than an *a priori* designer of choosing correctly

- Changing conditions can invalidate decisions
- Each application would have to adapt correctly

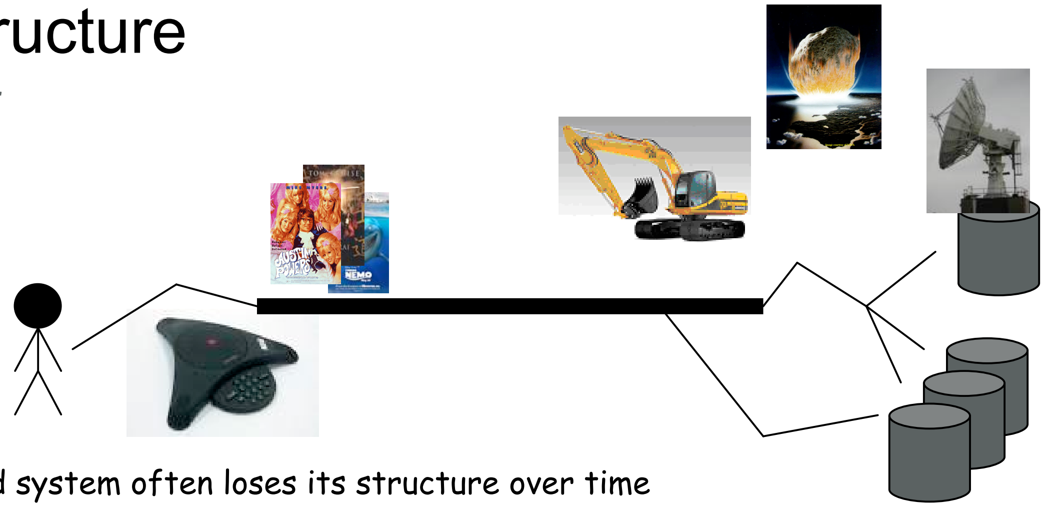
Bad idea 3: add `if` statements

People using arbitrary, unstructured, random conditionals should be hunted down

- Take a system that works, sometimes
- Look for some circumstances that call for a new optimisation
- Add an `if` statement (or equivalent) that detects these circumstances and applies the change in behaviour

The end of infrastructure

- We have to think of everything, in the small
- Everything may interact, so no composition



Unless you're careful, a rule-based system often loses its structure over time

Worst idea: assuming we know

All these ideas share a common problem, which we might term "the fallacy of the boolean assumption"

- There is a decision to be made
- …we collect the information we need and make it
- …and then we forget about it and move on

It neglects the uncertainty and dynamism that pervade autonomic and pervasive systems

- Any decision may be outdated by events
- Any decision may impact any other future decision
- We add (and hopefully remove) decision paths over time
- At the large scale, there is no "local"

…which are of course all straw men to some extent, although I got them all from the literature in some form or another

Ok, so give us some *good* ideas…

To find fault is easy; to do any better may be difficult.

– Plutarch

Might it be that our tendency to be reductionist is actually part of the problem?

- Divide the problem into small parts and fix them individually
- …which assumes both that there *are* small parts
- …*and* that individual fixes will compose to a global solution

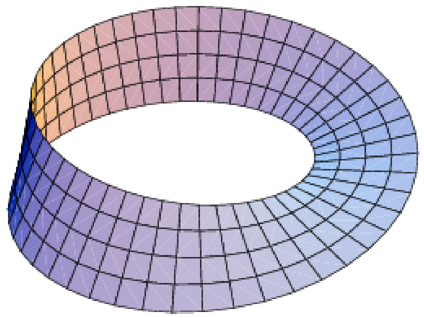
This isn't to argue for some kind of irreducible complexity, but simply to suggest that we need to reduce the problems along the right axes

- Capture and understand the system and the implications of its adaptations *as a whole*

Continuity

If the problems are arbitrariness and uncertainty, we need to find a model that protects us from them

- Systems go together in predictable ways, adding or removing sources of information are structured and well-defined processes
- Small changes due to uncertainty or errors do not cause enormous errors or wild swings in behaviour



The mathematics of continuity

- Topology, control theory, category theory, …
- Describe the system's responses to all variables as movement over a surface

Physicists call this approach a *phase space*

- A multi-dimensional picture of a system's behaviour
- Moving smoothly, avoiding unpleasant "jumps", is generally good

Discontinuity

Could we model autonomic behaviour in this way?

Not *all* changes of behaviour are continuous

- Hand-off to a different base station, for example

But we might try to *identify* the discontinuities, and then use various strategies to mitigate them

- The point isn't that behaviour is seamless, but rather that the seams have to appear in the right places
- Point correctness: do the right thing under the circumstances
- Process correctness: exhibit these correct behaviours correctly

Having a global model helps perform these analyses, and so identify the pitfalls

- May evolve over time, but maintain desired structure

Coutaz *et alia.* Context is key. Comm. ACM **48**(3). 2005.

Policy

Policies and other declarative forms

- Talk about goals rather than operations, map through to actions
- Declarative nature allows further analysis

Not necessarily easy to work with on a large scale, though there are promising strands of research

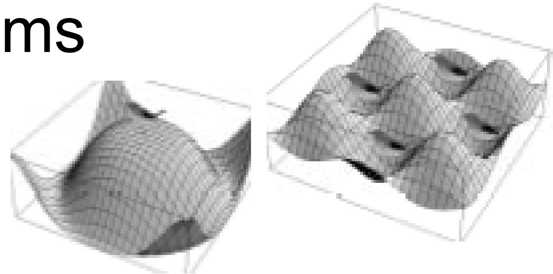
- Can we ensure separation of concerns?
- …and if not (as is probably the case), how do we detect the cross-over interference?
- Can we make the policies sufficiently noise-immune to deal with the uncertainties we know are there?
- Can we guarantee smooth behavioural changes?
- Can we do all this in the face of policy composition?

Self-stabilisation

We can avoid arbitrary decisions by embodying a continuous variation

For example, building an artificial "field" allows applications to mimic physical systems

- Each node contributes some information to the field
- Neighbours exchange "field strengths" and use them to affect their own local field
- Distortions propagate, damping encourages stability inherently from the core algorithm
- Well-defined variation and response to change



Mamei and Zambonelli. Programming stigmergic co-ordination with the TOTA middleware. ACM ICAA. 2005.

Similar results from using econometrics and game theory

Field analysis

Suppose we're getting incorrect information upon which we're relying to make decisions

- The local bandwidth is less than we're told, for example
- Continuity means that a small error will result in a small mis-adaptation
- Well-defined and simple – doesn't require complex decision-making
- Continuous – adapts proportionately to effects

Define an algebra of different fields, each dealing with a different aspect of the problem, and how they overlay (compose)

- Field composition defines a composite field
- Local rules define exactly how the composition works for a given collection of fields
- May get complex, but *locally* complex

Graph models

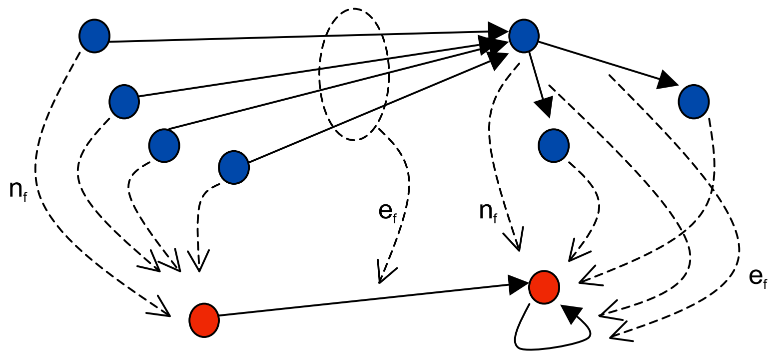
Getting more abstract, we model individual aspects of context and behaviour as graphs

- A context – the description of the sensed and inferred environment
- A situation – the behaviour or adaptation we should exhibit in response

Coutaz and Rey. Foundations for a theory of contextors. CADUI. 2002

Graph homomorphism $f = (n_f, e_f) : A \rightarrow B$

- $s_B(e_f(e)) = n_f(s_A(e))$ and similarly for edge targets
- Map nodes to nodes and edges to edges so as to preserve the adjacency structure
- So a context gives rise to a situation, which in turn defines the behaviour the network should exhibit



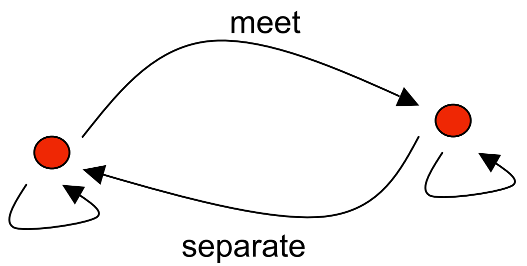
Location-based services

The system

- Two users a and b, being mapped to locations $l_1, l_2, \ldots, l_n$
- Gives rise to a context evolution graph with nodes the possible combinations of locations and edges the observations between them

We might want to exhibit different behaviour when the users are together than when they are apart

- Situation graph with two nodes, separate and meet edges, and identity loops

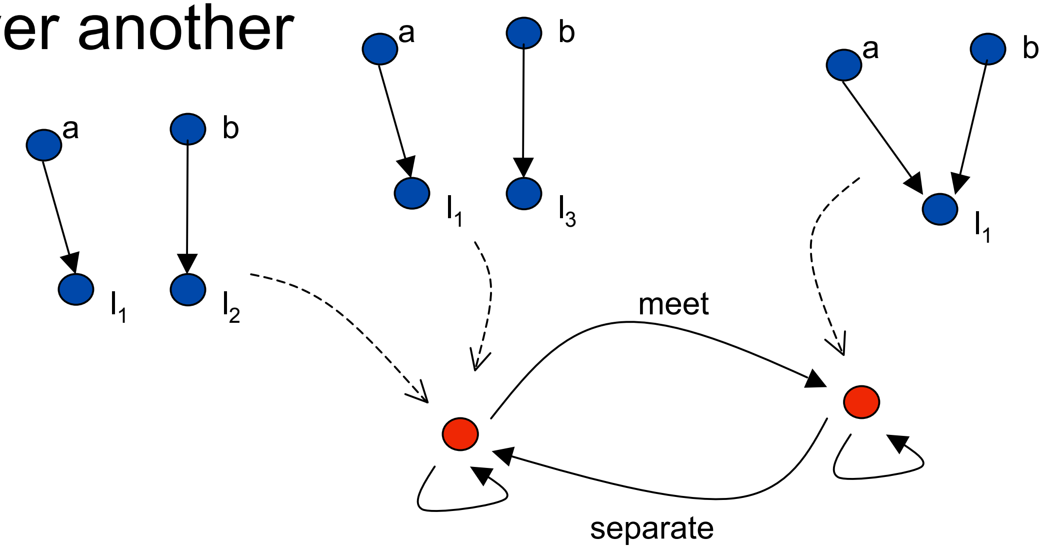


Dobson and Ye. Using fibrations for situation identification. Pervasive 2006 workshop proceedings. 2006.

Matching fibre to function

The contexts that give rise to each situation form a *fibration* of one graph over another

Boldi and Vigna. Fibrations of graphs. Discrete mathematics **243**. 2002



We can get too tied-up in adaptation – the *differences* – and forget the times we want to preserve aspects of the system – the *equalities*

Different contexts select the same adaptations, parameterised differently

- Match the environments to the behaviours, similarly with changes

Categorical descriptions

We can go a stage further, and use categories to describe our systems

- All information about context or behaviour becomes a category – a collection of objects and relationships with good compositional properties
- Very strong theory of composition, quite close to (functional) programming
- "Universal" descriptions of particular properties, so we can talk about strategies at a very abstract level which will apply to a great many situations

$$I \times I \xrightarrow{<e,f>} \underline{O}^o$$

$$\begin{array}{ccc} (O^I \times I) \times I & \xrightarrow{<ev,e>} & \underline{O} \\ \uparrow{\scriptstyle <(\overline{f'} \times I) \circ \pi_0, \pi_1>} & & \downarrow{\scriptstyle \pi_1} \\ C \times I & \xrightarrow[f']{} & O \end{array}$$

For example we can define what it means to be "the largest behaviour"

- System must not adapt to be "more general" than this
- Analyse compositions to make sure this is maintained

Envelopes of behaviour

What all these abstract descriptions give us is a level of mathematical tractability that we can use to predict and design the autonomic behaviours we want

- Analogous to what process algebra did for concurrency: useful notions of equality, characterised higher-level phenomena
- Can still be self-evolving, but within set bounds and goals
- Get what we want and exclude what we don't, in a well-founded manner
- Get complexity from composition

Dobson and Nixon. More principled design of pervasive computing systems. LNCS 3425. 2004.

All the schemes so far presented have flaws, but point in a similar direction

- Capture the general mathematical structures
- Let us deploy different formalisms within a common framework

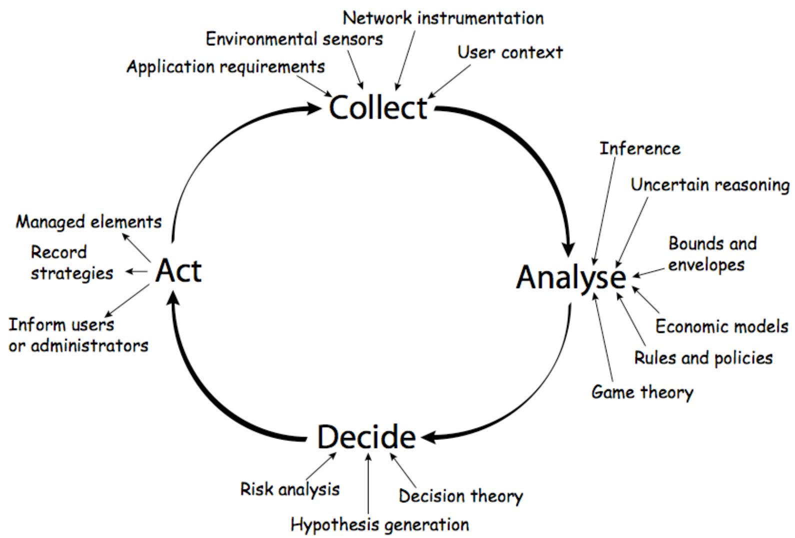
Towards a universal model

If we're to succeed on a large scale, we need a **universal theory of autonomic behaviour**

Realistic relationship with network technology, techniques and scales

Build these ideas directly into programming languages, so it's *easy* to express the systems we want – and *only* those systems

…but that's another talk :-)



An analytic framework that lets us predict what we'll get (and can therefore sell)

Open decision-making, able to combine the best techniques as they develop

Take-home messages

Six things to take away

1. Autonomics is multi-level: we need to account for information at **different semantic levels**
2. Our sensed information is **uncertain at a fundamental level**, and this must be reflected in our programming
3. In general, **any *ad hoc* approach will fail** when faced with complex context varying in complex ways
4. When interpreted widely, **continuity and compositionality** give us a structure for handling whole-system complexity
5. We can then **describe and predict a system's responses to stimuli** to maximise our ability to **engineer** autonomic systems
6. There's no obvious best way to do autonomic systems, so **universality and interoperability of analysis** are key

Conclusions

Autonomic networking can draw on a huge range of information, models and techniques

Our challenge overall is to continue to develop the exciting new approaches whilst retaining the measure of predictability and analysis networks require

All this work should of course be supported by large chunks of European and industrial money

And then we take over the world