

Challenges in very large distributed systems

Maarten van Steen
Guillaume Pierre, Spyros Voulgaris

VU Amsterdam, Dept. Informatics

vrije Universiteit *amsterdam*



Scenario

- Bob likes to discover music that he hasn't heard before fitting his rather diverse taste.
- For inspiration, he has subscribed to several music magazines, reading reviews and gradually knowing which reviewers he should pay attention to.
- Bob subsequently finds his way by buying CDs, hooking up to online stores, or streaming from online services such as Spotify.

In five years from now

- Bob likes to discover music that he hasn't heard before fitting his rather diverse taste.
- For inspiration, he has subscribed to an online music magazine, being offered reviews only from his favorite reviewers. Bob also writes reviews and submits these to the magazines.
- By tagging the music that seems interesting to him, his online music library is extended with music which he can later decide to purchase (according to different models).
- His audio system keeps track of when, where, and how often Bob listens to music, builds dependencies between musical fragments, and so on.

In five years from now

- Knowledge about Bob's listening habits is fed back into his electronic subscriptions, changing the priorities on what is to be offered.
- The system has discovered what the optimal search space is for Bob, including the “surprise” element.
- The recommendation system specifically takes into account the listening behavior from other people like Bob in order to optimize the search space for Bob (which includes reviews and favorite reviewers).

Note

You can fill in other stuff for “music,” like electrical appliances. Type of problems stay the same: **massive ubiquitous systems**.

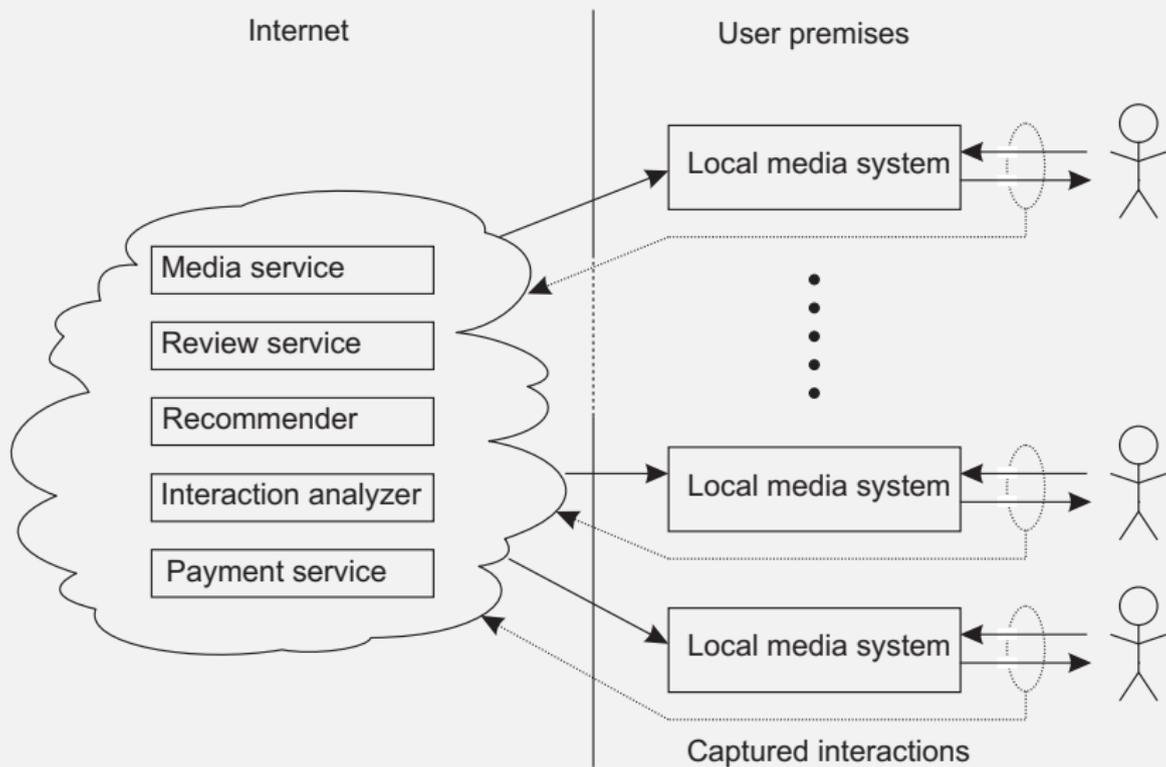
In five years from now

- Knowledge about Bob's listening habits is fed back into his electronic subscriptions, changing the priorities on what is to be offered.
- The system has discovered what the optimal search space is for Bob, including the "surprise" element.
- The recommendation system specifically takes into account the listening behavior from other people like Bob in order to optimize the search space for Bob (which includes reviews and favorite reviewers).

Note

You can fill in other stuff for "music," like electrical appliances. Type of problems stay the same: **massive ubiquitous systems**.

Overall system



Challenges: monitoring and analyzing user behavior

What's the issue?

We're dealing with millions of users for which we need to **nonintrusively** monitor behavior. The monitored behavior needs to be **combined** for further analysis.

Centralized versus decentralized?

Collecting data on a specific user should most likely be done locally, but moving massive amounts of data for aggregation with other observations may be out of the question \Rightarrow new workload distribution issues?

Challenges: monitoring and analyzing user behavior

What's the issue?

We're dealing with millions of users for which we need to **nonintrusively** monitor behavior. The monitored behavior needs to be **combined** for further analysis.

Centralized versus decentralized?

Collecting data on a specific user should most likely be done locally, but moving massive amounts of data for aggregation with other observations may be out of the question \Rightarrow **new workload distribution issues?**

Challenges: monitoring and analyzing user behavior

What's the issue?

We're dealing with millions of users for which we need to **nonintrusively** monitor behavior. The monitored behavior needs to be **combined** for further analysis.

Centralized versus decentralized?

Collecting data on a specific user should most likely be done locally, but moving massive amounts of data for aggregation with other observations may be out of the question \Rightarrow new workload distribution issues?

Challenges: making users part of the system

Observation

We are already witnessing that users are becoming computational resources: **crowdsourcing**

- Will Wikipedia be **personalized** – part of it becoming Bob's online music magazine?
- News becomes real news in Twitter because of the natural retweeting.
- There are games out there that have as a side effect that pictures are decently tagged.

Issue

Rethink the design of our very large online distributed systems not by considering **end users**, but considering **users as resources**. The distinction between system core and users is blurring.

Challenges: making users part of the system

Observation

We are already witnessing that users are becoming computational resources: **crowdsourcing**

- Will Wikipedia be **personalized** – part of it becoming Bob's online music magazine?
- News becomes real news in Twitter because of the natural retweeting.
- There are games out there that have as a side effect that pictures are decently tagged.

Issue

Rethink the design of our very large online distributed systems not by considering **end users**, but considering **users as resources**. The distinction between system core and users is blurring.

Challenges: making users part of the system

Observation

We are already witnessing that users are becoming computational resources: **crowdsourcing**

- Will Wikipedia be **personalized** – part of it becoming Bob's online music magazine?
- News becomes real news in Twitter because of the natural retweeting.
- There are games out there that have as a side effect that pictures are decently tagged.

Issue

Rethink the design of our very large online distributed systems not by considering **end users**, but considering **users as resources**. The distinction between system core and users is blurring.

Distributed systems are moving into the cloud

Observation

To make this all work, we need to collectively build, compose, run, and extend systems \Rightarrow call it the **cloud**.

Chaos all over the place?

- No single owner
- No single contributor
- No easy openness
- Can we organize matters:
 - What are the incentives to make crowdsourcing work?
 - Do we need to rethink application hosting (collaborative clouds?)
 - Can we facilitate **convergence** of solutions?

Let's manage less

Evolving systems

Design parameter spaces that can be evaluated and configured through **evolutionary algorithms**.

Example

Evolutionary clustering algorithms (decide on best parameters to cluster):

- Filename, artist, genre, album
- Parameters of the EA: to what extent allow crossovers and mutations?

In general...(and a bit vague)

Concentrate on trying to control only parts of emerging behavior.

Let's manage less

Evolving systems

Design parameter spaces that can be evaluated and configured through **evolutionary algorithms**.

Example

Evolutionary clustering algorithms (decide on best parameters to cluster):

- Filename, artist, genre, album
- Parameters of the EA: to what extent allow crossovers and mutations?

In general...(and a bit vague)

Concentrate on trying to control only parts of emerging behavior.

Let's manage less

Evolving systems

Design parameter spaces that can be evaluated and configured through **evolutionary algorithms**.

Example

Evolutionary clustering algorithms (decide on best parameters to cluster):

- Filename, artist, genre, album
- Parameters of the EA: to what extent allow crossovers and mutations?

In general...(and a bit vague)

Concentrate on trying to control only parts of emerging behavior.

Conclusions

Challenges

- Heading toward more and larger **socio-technical** systems, with users and system core in a tight loop.
- Fully decentralized and ad hoc composition needs to be facilitated.
- Impossible to fully monitor and control \Rightarrow design for managing less.