

Классификация веб-ориентированных систем публикования и подписки информационных ресурсов

Н. А. Алексеев, Л. С. Глоба, Marius Feldmann, Ю. Н. Молчанов
*Национальный технический университет Украины
«Киевский политехнический институт», Украина*

Complex analysis of publish/subscribe interaction paradigm was made, further developing prospects of systems based on such paradigm in modern Internet conditions were considered, also the classification based on systems functioning features was made. In this paper the additional classification based on system functioning features in global networks was proposed.

1. Актуальность темы

В настоящее время, стремительно увеличивается количество информации, ее потребителей и поставщиков в сети Internet [5]. Поскольку стандартные способы ее распространения и хранения становятся малоэффективными, существует потребность в использовании альтернативных способов, одним из которых является модель взаимодействия публикация/подписка. Она предоставляет свободную схему автономного взаимодействия участников информационного обмена, являясь мощным механизмом взаимодействия при работе в распределенных и мобильных архитектурах, позволяющим подписчикам выражать свой интерес в определенной информации, сгенерированной публикующим и впоследствии быть оповещенным о ее появлении.

2. Принцип взаимодействия публикация/подписка

Схема взаимодействия публикация/подписка, представленная на Рис. 1, состоит из трех участников: поставщика или публикующего, подписчика или потребителя и посредника.



Рис. 1. Объектно-ориентированная схема публикации/подписки

В отличие от классической модели взаимодействия запрос/ответ, модель взаимодействия публикация/подписка предоставляет три уровня разделения [1]:

1. Пространственное разделение представлено на Рис. 2 – участники взаимодействия не нуждаются в информации друг о друге. Публикация осуществляется через сервис событий, а подписчики получают эти события не напрямую, а через сервис событий. Публикующие обычно не имеют ссылку на подписчиков и не знают сколько подписчиков участвуют во взаимодействии. Также, подписчики обычно не поддерживают ссылки на публикующих.



Рис. 2. Пространственное разделение

2. Временное разделение представлено на Рис. 3 – участники взаимодействия не должны быть активны на протяжении взаимодействия одновременно. В частности, публикующий может опубликовать некоторые события в момент, когда подписчик отключен и наоборот.

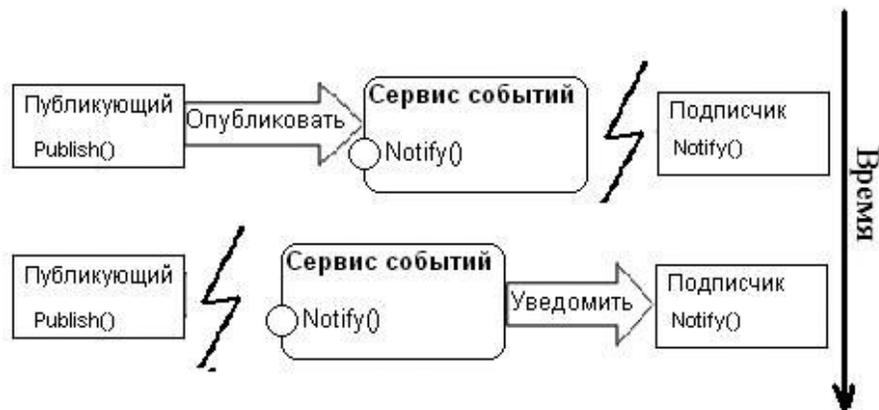


Рис. 3. Временное разделение

3. Разделение синхронизации представлено на Рис. 4 – публикующий может создавать событие в любое время, а подписчик может получать уведомление асинхронно (через обратный вызов).



Рис. 4. Разделение синхронизации

3. Существующая классификация систем публикации/подписки

Среди систем публикации/подписки (СП/П) можно выделить 3 типа, согласно [1]:

1. Темо-ориентированные – наследуют концепцию каналов, которые используются для связи коммуникационных узлов с методами классификации содержимого событий. Это означает, что поставщики могут публиковать события в соответствующих темах, а потребители подписываться на них. Темы идентифицируются по ключевым словам. Подписка на определенные темы означает участие в группе с названием данных тем. Соединение с темой требует выбора имени в формате URL. Темы организуются в дереве тем с помощью точечной нотации (например, wifi.радиосвязь.телекоммуникации) и подписчики могут подписаться на отдельную тему или на тему с ее подтемами. Каждая тема идентифицируется в сервисе событий уникальным именем. Когда в системе появляется событие, оно асинхронно пересылается всем подписчикам по этому URL. К таким системам относятся Corona (Cornell Online News Aggregator)[4], SCRIBE[2];

Основными достоинствами темо-ориентированных систем является простота реализации, малые задержки, высокая пропускная способность и простота переадресации подписок большому количеству пользователей, состоящих из групп через простое соответствие ключевых слов именам групп.

Среди недостатков можно выделить ограниченные выразительные способности семантики, невозможность подписки на часть темы и как следствие неэффективное использование пропускной способности.

2. Содержимо-ориентированные – уведомления о событиях сравниваются с хранимыми в базе данных подписками и соответствуют predetermined критерию в соответствии с содержимым событий, а не ключевым словом. Каждая подписка идентифицируется определением образца подписки, что необходимо для распространения события. Образец подписки является основой для определения фильтра подписки. Фильтры выбирают подходящие события из опубликованной информации с использованием языка подписки и реализуются в форме пар имя-значение свойств и базовых операторов сравнения. Образец подписки определяется одним из следующих способов: строки, объекты шаблонов, исполняемый код. Если система использует для определения образцов подписки строки, фильтры должны

согласовываться с семантикой подписки. Возможно использование SQL, XPath в качестве языков подписки.

Основным достоинством является повышение выразительности и, как следствие, повышенная гибкость механизма выбора оповещений.

Недостатками являются низкая производительность, сложная реализация масштабируемой системы. Наиболее характерным представителем данной группы СП/П является система SIENA[3].

3. Типо-ориентированные – оповещения делятся в соответствии с определенными типами, которые могут инкапсулировать атрибуты также как и методы. Подписчик указывает тип события и определяет выражение фильтра, которое оперирует с атрибутами, предоставляемыми этим типом события. Такой подход делает более тесной интеграцию языка программирования и промежуточного ПО. Фильтры содержимого определяются на родном языке, основанном на событиях, атрибутах и методах и могут быть применены удаленно для уменьшения загрузки сети. Подписчик определенного типа объектов получает только экземпляры класса такого типа и его подтипов.

Главным достоинством таких систем является простота и гибкость распределенной реализации посредством использования удаленной фильтрации содержимого, что позволяет увеличить количество возможных клиентов и трафика.

4. Предлагаемая классификация

Рассмотренная выше классификация может быть использована для описания СП/П с точки зрения их предназначения, задач, выполняемыми системами. Однако для описания СП/П с точки зрения особенностей функционирования систем в глобальных телекоммуникационных сетях целесообразно использовать критерии:

1. Используемого промежуточного ПО.
2. Используемых сетевых протоколов.
3. Источников подписки.

Согласно предлагаемой классификации можно выделить такие категории систем:

1. Системы, не требующие разработки и использования дополнительной клиентской или серверной части, например дополнительных модулей браузера или сервера, сетевых сервисов и протоколов, например HTTP, TCP/IP, WebDAV, SOAP, NNTP, позволяющие осуществлять подписку на разные источники, например БД, службы новостей, RSS.
2. Системы, требующие дополнительное ПО, требующие специальных протоколов с изменяемым источником.
3. Системы, требующие дополнительное ПО, не требующие специальных протоколов с фиксированным источником.
4. Системы, требующие дополнительное ПО, не требующие специальных протоколов с изменяемым источником. К таким системам относятся Corona, Webvigil.

5. Системы, требующие дополнительное ПО, требующие специальных протоколов с фиксированным источником.
6. Системы, не требующие дополнительное ПО, требующие специальных протоколов с фиксированным источником. К таким системам можно отнести системы мультимедийного видео.
7. Системы, не требующие дополнительное ПО, требующие специальных протоколов с изменяемым источником.
8. Системы, не требующие дополнительное ПО, не требующие специальных протоколов с изменяемым источником.

7. Выводы по результатам и направления дальнейших исследований

Предложенная классификация сможет упростить выбор типа системы публикации/подписки как исходя из сложности ее внедрения и функциональных возможностей, что важно для внедряющей ее организации, так и исходя из используемых протоколов и промежуточного ПО, что важно для разработчика при планировании архитектуры.

Анализируя возможности и недостатки каждого типа систем публикации/подписки можно сказать, что при использовании производительного алгоритма детектирования изменений в древовидных структурах, а также объектно-ориентированного языка программирования, позволяющего увеличить производительность и простоту реализации, наиболее целесообразным для сети Internet является разработка именно содержимо-ориентированной системы, которая бы относилась к системам, не требующим дополнительного ПО, не требующим специальных протоколов, обладающим возможностью смены источника подписки. Такая система сможет использовать стандартный инструментарий, существовавший как до появления группы технологий Web 2.0, так и последующие версии платформы, а именно работать по протоколу HTTP без использования промежуточного ПО, осуществлять подписку на изменение на выбранном пользователем Web ресурсе, используя технологию XPath. Это облегчит ее интеграцию и использование подписчиками, даст возможность работать как с документами XML, так и HTML, которые являются частным случаем XML.

ЛИТЕРАТУРА

1. Patrick Th. Eugster, Pascal A. Felber, Rachid Guerraoui, Anne-Marie Kermarrec. The many faces of publish/subscribe // ACM Computing Surveys (CSUR). – 2003. – v. 35, № 2. – P.92-107.
2. Castro M., Druschel P., Kermarrec A., Rowstron A. 2002. SCRIBE: A large-scale and decentralized application-level multicast infrastructure // IEEE J. Sel. Areas Commun. – 2002. – м.20, №8. – P. 1489-1499.
3. Carzaniga A., Rosenblum D., Wolf A. Achieving scalability and expressiveness in an Internet-scale event notification service // In Proceedings of the Nineteenth

- ACM Symposium on Principles of Distributed Computing (PODC '00). – 2000. – v.1, №1. – P.21-28.
4. Ramasubramanian V., Peterson R., GunSiree E. Corona: A High Performance Publish-Subscribe System for the World Wide Web // NSDI.-2006.-P. 15–28.
 5. Общие объемы информации превысят 1800 экзабайт в 2011 [chip.com.ua] <http://chip.com.ua/63000.html>