Towards the Future Internet
G. Tselentis et al. (Eds.)
IOS Press, 2009

doi:10.3233/978-1-60750-007-0-102

The Publish/Subscribe Internet Routing Paradigm (PSIRP): Designing the Future Internet Architecture

Sasu Tarkoma, Mark Ain, Kari Visala
Helsinki Institute for Information Technology (HIIT), P.O. Box 9800, 02015 TKK, Finland
{Sasu.Tarkoma, Mark.Ain, Kari.Visala }@hiit.fi

Abstract. Despite its success, the Internet is suffering from several key design limitations, most notably the unification of endpoint locators and identifiers, and an imbalance of power in favor of the sender of information. The unfavourable consequences are that the full range of possibilities offered by the Internet may not be fully realized and trust in its proper operation has been significantly weakened. In this paper, we introduce the Publish/Subscribe Internet Routing Paradigm (PSIRP) and present an architectural redesign of the global Internet based on an information-centric publish/subscribe (pub/sub) communication model. Through its application of pub/sub communications and efficient network design emphasizing end-to-end trust, we believe that the PSIRP-reengineered Internet may resolve many of the problems plaguing the current Internet and provide a powerful and flexible network infrastructure with a high degree of resiliency.

Keywords: Future Internet, information-centric, publish/subscribe, scoping

1 Introduction

Since its conception, the Internet has experienced tremendous growth, ever increasing traffic and new applications, including voice and video, while still retaining its original architecture drafted almost 40 years ago. The main guiding principle for the design of the Internet was the end-to-end principle [Sal1984].

Blumenthal et al. [Blu2001] identify a number of challenges for the end-to-end principle: operation in an untrustworthy Internet, more demanding applications, the rise of third party involvement, ISP service differentiation, and less sophisticated users. Moreover, one of the most notable issues in the current Internet is the imbalance of powers in favor of the sender of information, who is overly trusted. The network accepts anything that the sender wants to send and will make a best effort to deliver it to the receiver. This has led to increasing problems with unsolicited traffic (e.g. spam e-mail) and distributed denial of service (DDoS) attacks.

The *publish/subscribe (pub/sub) paradigm* has been proposed as a remedy to the problems facing the current Internet. In pub/sub networking, senders “publish” what they want to send and receivers “subscribe” to the publications that they want to receive. In principle, no one receives any material to which they have not explicitly expressed an interest by way of subscription.

One can observe that many widely used Internet applications already are essentially publish/subscribe in nature. For example, distribution of software updates is currently performed in a poll/unicast fashion that is clearly non-optimal. Instead, subscribing to the updates that are needed and distributing them via multicast, caching etc. would be much easier and more efficient from the point of view of network resource usage.

The PSIRP project will redesign the entire Internet architecture from the pub/sub point of view, taking nothing (not even IP) for granted. PSIRP's work will focus on the intersection of security, routing, wireless access, architecture design, and network economics, in order to design and develop efficient and effective solutions. In such a new Internet, *multicast* and *caching* will replace unicast and cache-free data fetching operations, while *security* and *mobility* will be embodied directly into the foundation of the architecture rather than added as after-thoughts.

The new pub/sub-based internetworking architecture aims to restore the balance of network economics incentives between the sender and the receiver and is well suited to meet the challenges of future information-centric applications and use modes. To our knowledge, this type of application of pub/sub communication models has not been tried before.

This paper is structured as follows: Section 2 provides an overview of the concept of information-centric networking and general philosophy behind PSIRP; Section 3 covers the architectural components, entities, processes, network services, and their functionalities; Section 4 discusses the prototype implementation and application considerations, and finally Section 5 contains concluding remarks.

2 PSIRP Background

We aspire to change the routing and forwarding fabric of the global inter-network so as to operate entirely based on the notion of information (associated with a notion of *labels* to support fabric operation) and its surrounding concerns, explicitly defining the *scope* of the information and directly addressing information (via *rendezvous identifiers*) as opposed to addressing physical network endpoints. The envisioned operation on information is in sharp contrast to the current endpoint-centric networking model. The current end-to-end model of IP networking requires that both the relevant data and explicitly-addressed network locations be known in order to transparently stream information between two endpoints. Our model emphasizes information-centric operation: data pieces are explicitly addressed through identifiers and labels serving as high-level designations/resolvers to lower-level schemas, and scoping mechanisms that can define information inter-networks and relationships within a global information taxonomy. As such, information is embedded immediately into the network and it is the only effective element in need of direct user-manipulation; the physicality of the network (i.e. endpoint locations) need not be known directly.

Another important aspect of the PSIRP architecture is that it is receiver-driven. We take the approach that the receiver has control over what it receives and we cascade this approach throughout the core of the PSIRP *component wheel* and the multiple operational elements within the PSIRP architecture. A receiver must *elect* to join (i.e., subscribe) to an identifier before it can receive any information. Sending (i.e., publishing) as well as receiving operations are thus decoupled between the senders and the receivers in both space and time. Hence, PSIRP not only intends to move the functionality of many existing publish/subscribe systems (e.g., [Eug2003b]) onto the internetworking layer but also base the entire communication, throughout the architecture, on this paradigm.

3 PSIRP Conceptual Architecture

This section presents the PSIRP conceptual architecture, defining the key entities and processes of system and their attributes. The PSIRP conceptual architecture consists of three crucial parts, namely the protocol suite architecture (called the component wheel), the networking architecture, and the service model.

3.1 Component Wheel

The PSIRP conceptual architecture is based on a modular and extensible core, called the *PSIRP component wheel*. The architecture does not have the traditional stack or layering of telecommunications systems, but rather components that may be decoupled in space, time, and context. The idea of such a layer-less network stack has been proposed before, for example, in the Haggle architecture [Hag2007]. The novelty of the PSIRP proposal is to use publish/subscribe style interaction throughout the conceptual architecture, and thus support a layer-less and modular protocol organization. This organization is primarily achieved through the efficient structuring of information identifiers and their interactions amongst network elements, offering ample flexibility for future expansion.

Figure 1 presents an outline of the conceptual architecture with the PSIRP component wheel in the middle. Above the wheel, we have APIs that facilitate accessibility to and implementation of different networking features that are available in the system. The figure illustrates the typical components needed in the wheel for inter-domain operation: *forwarding, routing, rendezvous,* and *caching*.

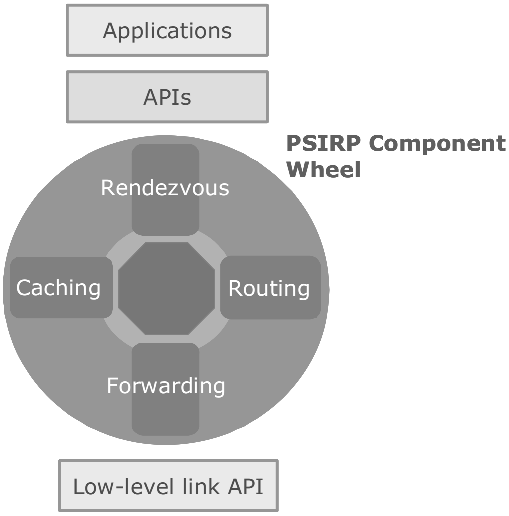

Fig. 1. PSIRP component wheel.

3.2 Network Architecture

We can view the global network of information as an *acyclic graph* of related pieces of data, each identified and scoped by some identifiers. In the PSIRP architecture, identifiers define the relationships between the pieces of information in our system on the different levels, such as the application or networking level. With this in mind, we propose the following classes of identifiers:

- *Application identifiers (AId),* used directly by publishers and subscribers.
- *Rendezvous identifiers (RId)*, used to bridge higher level identifiers with lower layer identifiers.
- *Scope identifiers (SId)*, used to delimit the reachability of given information.
- *Forwarding identifiers (FId)*, used to define network transit paths and transport publications across networks.

A rendezvous identifier is implicitly associated with a well-defined (but not necessarily fixed) *data set*, consisting of one or more publications. The data sets may also have associated *metadata*, which may include, e.g., scoping information and other useful information either for ultimate receivers or for network elements.

We also consider metadata that is understood within the network itself. Such network-level metadata might be concerned with how the communication for a particular data item may be conducted. This network metadata may be found as soft state within the network, carried as part of the communication header, or referred to by separate identifiers. Such functions may include access control, flow control, error notification, congestion notification, etc.

PSIRP necessitates the existence of a *scoping* mechanism as to limit the reachability of information. Scoping information is associated with a publication, determining the elements of the rendezvous system that act on published data and therefore defines the information network that the information belongs to. A publication may be associated with one or more scopes.

Scoping can be seen to represent a logical equivalent to the concept of *topologies* (such as link-local or site-local) in the endpoint-centric IP world. Given the information-centrism of our architecture, however, a scope is naturally attached to every item of information that is fed into the network (although we can consider the case of "no scope" being attached to a data item as being equivalent to the notion of "localhost" in the IP world - in other words, the information would not leave the local computer). In effect, scoping allows for building *information inter-networks* as opposed to topological inter-networks.

Scopes define a powerful concept that can construct social relations between entities, representing consumers and providers of information, and the information itself. This is illustrated in Figure 2, where certain information (e.g. a picture) is available to family and friends, while other information is merely visible to colleagues. Each scope is attached with a governance policy, represented as metadata, which may include (amongst other things) authentication information for potential receivers of the information.

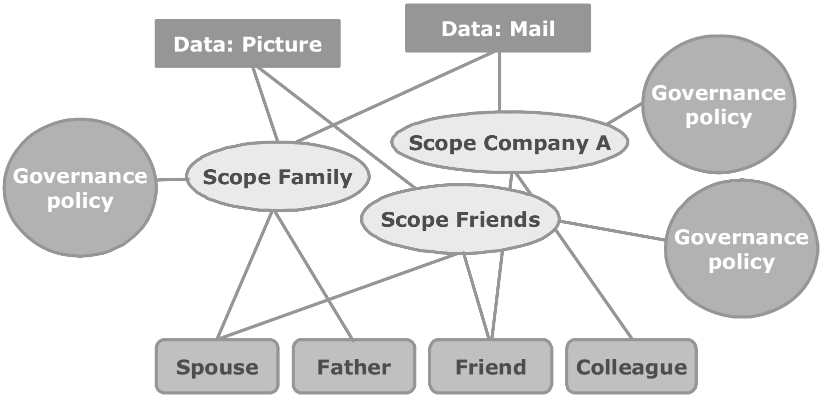

Fig. 2. Concept of scope.

Scopes can be easily re-constructed, removing certain parties from the scope, adding new publications to the scope, and assigning information to a new scope. This ability works towards our stated ambition to enable networks to reflect social structures and information grouping in a dynamic fashion.

The publisher/sender interface supports publication of data. Each publication has an associated flat label (i.e. the rendezvous identifier), and an optional metadata part.

A subscriber initiates a receiver-driven communication through a rendezvous identifier, specified in an act of subscription. Similar to the publisher, the subscriber can specify additional metadata surrounding the request.

3.3 Functional Entity Relationships

Figure 3 illustrates the relationships between the key entities of the PSIRP architecture.

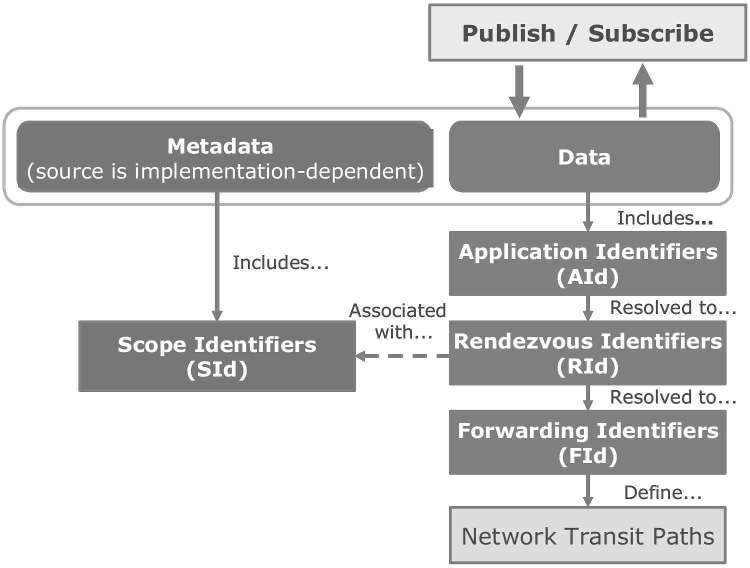

Fig. 3. Key entities of the conceptual architecture.

Typically, data is associated with one or more application identifiers and one or more scopes. Each application first resolves *application identifiers (AId)* into rendezvous identifiers.

A *rendezvous identifier (RId)* represents the network level identity of a publication and is associated with policy-compliant data dissemination graphs for publication delivery, both in the local domain (intra-domain) and between domains (inter-domain). The rendezvous identifiers are chosen from within a large enough set to provide a probabilistically feasible likelihood of uniqueness without a central allocation authority

A given application then hands the rendezvous identifiers to the network, using the scopes to properly map each rendezvous identifier to one or more *forwarding identifiers (FId)*, both within a domain and between domains. It is then the responsibility of the rendezvous functions to find suitable data transit and delivery paths in the network and denote them with forwarding identifiers. The breadth of reference of FIds is variable, potentially limited to single hops or dynamically expandable to encompass full multicast trees. This relatively open structuring scheme allows concurrent use of FIds to support flexible routing mechanisms based on source routing, anycast, multicast trees etc.

3.4 Rendezvous and Routing

Rendezvous is the process of resolving rendezvous identifiers into forwarding identifiers within a given scope. The scope determines the part of the rendezvous system that is used by the network. The three simplistic, topology-oriented cases, reflecting the current usage, are *link-local*, *intra-domain*, and *inter-domain* scopes. The multiple scopes of rendezvous are depicted in Figure 4. However, we expect that future applications will use more semantically-based scopes, implementing, e.g., scopes based on social networking structures.

Due to its importance in policy enforcement and defining (often user-created) information scopes in various situations, the rendezvous system constitutes a relatively well-defined environment where tussle is likely to commence [Cla2002]. The rendezvous system is therefore a policy-enforcement point in the architecture and a mechanism for supporting freedom of choice for network end points. Similar rendezvous functionality has been used in many distributed systems, for example the HIP [Mos2008] [Egg2004], IP multicast [Dee1998], i3 [Sto2002] and Hi3 [Nik2004], FARA [Cla2003], PASTRY [Row2001], and HERMES [Pie2004].

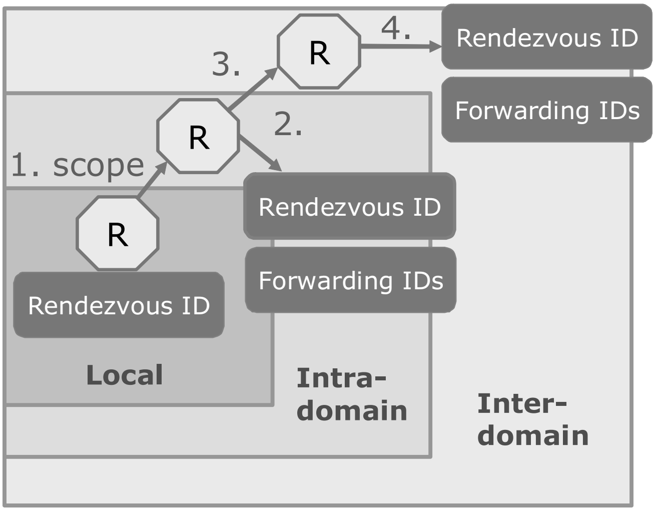

Fig. 4. Network architecture with rendezvous.

Rendezvous state is created to allow the subscriptions and publications to meet within a specified scope. Subscriptions and pre-publication notifications (or *advertisements*), possibly included in metadata or data-correlation notifications, may be utilised by the rendezvous system to create partial forwarding state from publishers towards subscribers. When a publication becomes active, i.e., when there is both an actively sending publisher and one or more active subscribers, the rendezvous systems are used to complete the forwarding path by mapping the rendezvous identifier to intra-domain and inter-domain forwarding identifiers. This late mapping can be used to implement dynamic policies, such as inter-domain routing and caching.

The rendezvous system ensures that neither traffic policies nor publication/subscription policies and scopes are violated. We will cover the efficiency, policy, and incentive issues related to the inter-domain rendezvous in detail in our forthcoming publication. Replication is used as the main method to achieve resilience against failures in the rendezvous system.

Routing processes in PSIRP are categorized as either intra-domain or inter-domain. Intra-domain routing is concerned with delivery within an administrative domain. Inter-domain routing pertains to data delivery in the global network, typically spanning several domains. FIds specify the policy compliant paths on the level of domains which makes them more tolerant of router failures along the path.

A subset of the forwarding routers may store cached copies of publications for faster access and time-decoupled multicast lessening the burden of the publisher. The publications are persistently stored by the publishing nodes and the network state can be considered to be fully a soft state that can be recovered in the case of a failure.

Both the rendezvous process and the payload forwarding can be protected by cryptographic means to provide integrity and authenticity of information on packet level as described in [Lag2008].

4 Prototype Implementation

The implementation work has focused on an intra-domain implementation of the PSIRP architecture based on Ethernet. The modular prototype implementation structure is illustrated by Figure 5. The implementation and experimentation work is currently on-going.

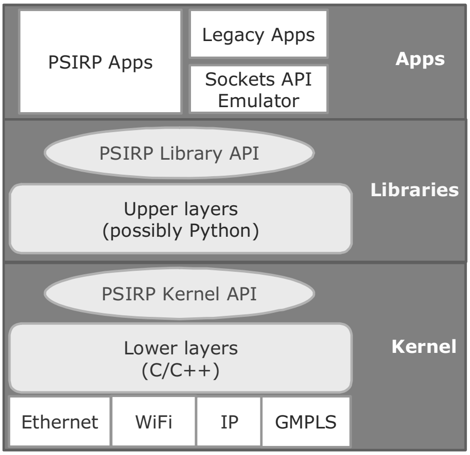

Fig. 5. Implementation structure of the PSIRP prototype.

The core PSIRP prototype will be implemented in two layers:

1. The lower layer will mainly reside in kernel space for performance reasons, providing whatever functionality is deemed to be critical enough to justify its inclusion into the kernel. Lower-level protocols such as Ethernet, Wifi, and IP are included for legacy support; GMPLS offers advanced label-switching

and traffic engineering functions that positively compliment PSIRP's technical ambitions.

2. The upper layer(s) will reside exclusively in user space for reasons of flexibility, providing whatever functionality is deemed necessary to support writing PSIRP applications, supplementing the functionality provided by the lower layer.

5 Conclusions

This paper presents the first results of the architectural design process within the PSIRP project. It can only be seen as the first step in the desired clean slate redesign of the Internet. We envision a process of bottom-up lessons learned and top-down rationalization, the first results of which are visible in this report.

Following this methodology, the conceptual architecture and components presented in this paper are only part of our progress so far in the project. The clarification of concepts, presented in the design considerations, as well as the formulation of new questions pushing forward our future development, are central to the background work we have made. Hence, many of the issues addressed in this paper, such as identifiers, the concept of scope, rendezvous, caching, forwarding and transport as well as our inter-domain routing solution, will see further progress in the near future.

Acknowledgements

This work has been funded by the EU FP7 PSIRP project (Contract INFSO-ICT-216173). The paper is based on the work of the PSIRP architecture and implementation teams, including Pekka Nikander (LMF), Dirk Trossen (BT), Trevor Burbridge (BT), András Zahemszky (ETH), Jarno Rajahalme (NSN), Dmitrij Lagutin (TKK-HIIT), Mikko Särelä (LMF), Janne Riihijärvi (RWTH), and Teemu Rinta-aho (LMF).

References

[Ber2001] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American Magazine*, May 17, 2001, available at: http://www.sciam.com/article.cfm?id=the-semantic web [Accessed on 15 July, 2008].

[Blu2001] M. Blumenthal and D. Clark, "Rethinking the design of the Internet: The End-to-End arguments vs. The Brave New World," *ACM Transactions on Internet Technology 2001*, vol. 1, issue 1, 2001, pp. 70-109.

[Cla2002] D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden, "Tussle in Cyberspace: Defining Tomorrow's Internet," in *Proc. ACM SIGCOMM Conference on*

Applications, Technologies, Architectures, and Protocols for Computer Communications, New York, NY, Aug. 2002, pp. 347-356.

[Cla2003] D. Clark, R. Braden, A. Falk, and V. Pingali, "FARA: Reorganizing the Addressing Architecture," in *Proc. ACM SIGCOMM Workshop on Future Directions in Network Architecture*, Karlsruhe, Germany, Aug. 2003, pp. 313-321.

[Dee1998] S. Deering, D. Estrin, D. Farinacci, M. Handley, A. Helmy, V. Jacobson, C. Liu, P. Sharma, D. Thaler, and L. Wei, *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification*, IETF RFC 5201, June 1998.

[Egg2004] L. Eggert and J. Laganier, *Host Identity Protocol (HIP) Rendezvous Extension*, IETF RFC 5204, April. 2008.

[Eug2003b] P. T. Eugster, P. A. Felber, R. Guerraoui, and A. M. Kermarrec, "The Many Faces of Publish/Subscribe," *ACM Computing Surveys (CSUR)*, vol. 35, issue 2, pp. 114-131, 2003.

[Hag2007] Haggle partners, "Deliverable D1.2: Specification of the CHILD Haggle," *HAGGLE*, Aug. 2007, available at: http://www.haggleproject.org/deliverables/D1.2_final.pdf [Accessed on 15 July, 2008].

[Jac2006] V. Jacobson, "If a Clean Slate is the Solution What Was the Problem?," *Stanford "Clean Slate" Seminar*, Feb. 2006, available at: http://cleanslate.stanford.edu/seminars/jacobson.pdf [Accessed on 15 July, 2008].

[Kat2006] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in The Air: Practical Wireless Network Coding," ACM SIGCOMM Computer Communication Review, vol. 36, issue 4, pp. 243-254, 2006.

[Lag2008] D. Lagutin, "Redesigning Internet - The packet level authentication architecture," licentiate's thesis, Helsinki University of Technology, Finland, June 2008.

[Mos2008] R. Moskowitz, P. Nikander, P. Jokela, and T. Henderson, *Host Identity Protocol*, IETF RFC 2362, Apr. 2008.

[Nik2004] P. Nikander, J. Arkko, and B. Ohlman, "Host Identity Indirection Infrastructure (Hi3)", in *Proc. Second Swedish National Computer Networking Workshop (SNCNW 2004)*, Karlstad, Sweden, 2004.

[Per2002] A. Perrig, R. Canetti, J. D. Tygar, and D. Song, "The TESLA Broadcast Authentication Protocol," CryptoBytes, vol. 5, no. 2, Summer/Fall 2002, pp. 2-13.

[Pie2004] P. R. Pietzuch, "Hermes: A Scalable Event-Based Middleware," doctoral dissertation, Computer Laboratory, Queens' College, University of Cambridge, Feb. 2004.

[Psi2008] M. Ain, S. Tarkoma, D. Trossen, P. Nikander (eds.), "PSIRP Deliverable D2.2: Conceptual Architecture of PSIRP Including Subcomponent Descriptions", available at http://www.psirp.org/.

[Psi2008b] D. Trossen (ed.), "PSIRP Vision Document", available at http://www.psirp.org/ [Accessed on 15 July, 2008].

[Row2001] A. Rowstron and P. Druschel, "Pastry: Scalable, Decentralized Object Location and Routing for Large-Scale Peer-to-Peer Systems," in *Proc. of Middleware 2001*, Heidelberg, Germany, Nov. 2001, pp. 329–250.

[Sal1984] J. H. Saltzer, D. P. Reed, and D. D. Clark, "End-to-End Arguments in System Design," *ACM Transactions on Computer Systems*, vol. 2, issue 4, pp. 277-288, Nov. 1984.

[Sto2002] I. Stoica, D. Adkins, S. Zhuang, S. Shenker, and S. Surana, "Internet indirection infrastructure," in *Proc. 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, Pittsburgh, PA, Aug. 2002, pp. 73-86.