Peer-to-Peer XML

Presented By:
Vasu Sathu
11/11/02

1

Overview

- Introduction to P2P
- XML and P2P Technology
- Applications of P2P
- Security Issues in P2P
- Future Aspects
- Conclusion

2

Introduction to P2P

- **Definition**:

"P2P is about unifying the roles of Client and Server".

" P2P is a class of applications that takes advantage of resources, storage, cycles, content and human presence available at the periphery of the Internet".

- Clay Shirky

3

Introduction to P2P

- Motivation to Adopt P2P :
  - Decentralized Network Architecture.
  - Cost and Effective Resource Allocation.
    - Proper utilization of storage, bandwidth and computing resources.
  - Pervasive Computing
    - Edge services acts as network caching mechanisms.

4

Introduction to P2P

- Exchange of information is done between peers with minimal involvement of centrally managed servers.
- No concept of client and server.
- Every peer is active participant and has equal status.
- Administration must be done at every peer leading to overhead in enforcement of security policies.

5

Introduction to P2P

- P2P paradigm is classified in to two major types.
- Hybrid P2P –The central server is responsible for maintaining a catalog of addresses referenced by a set of indexes.

  Ex: Napster.
- Pure P2P – Every node is a peer having same equal capability with no central servers.

  Ex: Gnutella , Freenet.

6

Introduction to P2P

- Characteristics of P2P
  - No Central Co-ordination.
  - No Central Database.
  - No peer has global view of the system.
  - Global behavior emerges from local interactions.
  - Peers are Autonomous.
  - Peer connections are unreliable.

7

Introduction to P2P

- Operations of P2P
  - Identity – Name and credentials of a peer.
  - Discovery – Finding what services and resources are currently available.
  - Authentication – Securely verifying the identity of a peer.
  - Authorization – Access privileges to perform a given task.

8

XML and P2P Technology

- P2P systems have no central servers for dispatching information.
- Designing a mechanism for peers to communicate is a critical aspect.
- Data resides locally and efficiently distributing data is not a trivial task.
- Managing the updates to the peer application is cumbersome.
- P2P systems that embrace XML can provide the basis of a solution.

9

XML and P2P Technology

- The areas in P2P technology that benefit most from XML are :
  - Messaging
  - Content Management
  - Data Storage
  - Application Deployment
  - Profiling

10

XML and P2P Technology

- **Messaging**
  - XML offers an effective solution for sending short structured messages for communication among peers.
  - XML messages allows different kinds of peers to discover,participate and communicate with each other independent of their protocols.
  - P2P applications such as Napster and Gnutella have their own protocols for communication and by using XML as the underlying protocol these applications can interact with each other.

11

XML and P2P Technology

- XML messages can be easily generated and parsed by applications.
- XML data can be encrypted making it suitable candidate for secure messaging.
- Applications like XML- RPC and Jabber have influenced development of P2P applications.
- Hence different types of applications that work on different platforms can interact and exchange information.

12

XML and P2P Technology

- **Content Management**
  - In E-Commerce environment corporations having business have to exchange information with partners, clients, customers etc. It can be done in two ways :
    - **Web Portals**
    - **Replicated Servers**
  - Web portals are centralized stores of data where all the users share and update information.
  - This is not only a lot of work and cumbersome but disruptive to work processes that require data remain in its original locations.

13

XML and P2P Technology

- In Replicated Servers each department uses its own server for storing information and they will be updated with each of the other data stores at regular intervals of time.
- Disadvantage is user cannot get the updated data since data stores are updated at particular intervals.
- P2P networks solve this problem by allowing access to decentralized content since the peers can access and provide information in real time.
- P2P networks use XML for communicating and efficiently accessing the decentralized information.

14

XML and P2P Technology

- **Data Storage**
  - Data resides locally on the peer for processing.
  - Locally caching application data in P2P systems provides flexibility.
  - Data retrieval is efficient with structured data.
  - XML handlers can be used to search ,validate and manipulate data needed to support peer application and improves application performance.

15

XML and P2P Technology

- **Application Deployment**
  - As the peers in a P2P network expands, issues with application deployment increases.
  - An XML based solution is Open Software Description (OSD) to solve the software updates.
  - OSD allows system architects to define the components required for peer applications along with the download location and component dependencies.
  - Every peer can verify whether it has the most recent software and automatically download upgrades.

16

XML and P2P Technology

- Integrating OSD files in to P2P deployment strategy shifts the burden from user to P2P application itself.

- **Profiling**
  - Without access to central services, XML data files will be used to customize applications to user preferences and security.

17

Applications of P2P

- Case study (Jabber)
  - An open XML and fully distributed protocol for the real time exchange of messages between peers.
  - It is aware of the XML namespaces which permits different groups of people to define different sets of XML to present data.
  - Instant messaging using Jabber can combine two different namespaces to exchange information.
  - Jabber enables the inclusion of any XML data in any namespace thus allowing applications to include, intercept and modify their own XML data.

18

Case Study : Jabber

- Here is a simple message using Jabber's XML format

```
<message to="sam@capulet.com" from=bill@montague.net"
 type="chat">
  <body> Here, sweet lord, at your service. </body>
  </message>
  And here's a hypothetical message with additional data in the
     namespace included:
 <message from="tom@capulet.com" to=bill@montague.net">
    <body>Angels and Ministers of Grace, defend us! </body>
    <prayer xmlns="http://www.grace.org">
    <verse>.....</verse>
    </prayer>
  </message>
```

19

Case Study : Jabber

- Jabber architecture closely resembles "email".
- The Jabber server plays three important roles
  - Handling client connections and communicating directly with Jabber clients
  - Communicating with other Jabber servers.
  - Coordinating the various components associated with the server.
- Jabber ID (or JID) contains a set of ordered elements. The JIDs are formed of a domain, node and the resource as in the format

  [node@/domain[/resources]

20

Case Study : Jabber

- Domain name is the primary identifier and represents the Jabber server to which the entity connects.
- Node is the secondary identifier and represents the user within a specific domain.
- Resource belongs to a node and is used to identify specific objects that belong to a user.
- A Jabber user always connects to the server by means of a particular resource and therefore has address of the form node@domain/resource.

21

Security Issues

- Security concern is one of the primary issues in P2P.
- P2P applications can be broadly classified in to three categories :
  - File Sharing – Since peers share data they might get viruses through downloads.
  - Another problem is exploitation of installed client software bugs so that the hackers can utilize bugs in client software to gain control of the system.
  - Instant Messaging – There may be leakage of information among peers which can be exploited by other peers

22

Security Issues

- A solution to this problem is to encrypt the messages sent to each peer.
  - Distributed Computing – Large complex applications are distributed among peers for computational analysis and resources.
    - A security risk is Trojan horse problem which collects data from your system and sends them to a hacker.
- Intel provided a security framework for the creation of secure P2P applications

23

Security Issues

- PTPTL(Peer to Peer Trust Library) is a good choice for developing secure P2P applications
- PTPTL is built upon the OpenSSL Toolkit and its a open source framework.
- PTPTL supports security issues such as digital certificates , peer authentication, secure storage, private key encryption and digital signatures.
- Advantage of using PTPTL is that its open source, no license fees for using the application and its based on the existing security standards such as X.509.

24

Future Aspects

- Evidence of strong interest in P2P technologies.
- Both Gnutella and Freenet have attracted significant number of users.
- Sun's JXTA has also gained increasing popularity.
- Future Operating systems will incorporate P2P functionalities such as security, file transfer and messaging.
- Convergence of P2P with Web Services.

25

Conclusion

- This technology is still evolving
- It will find its application in diverse areas, although it started with mere file swapping.
- P2P networking technologies can greatly improve the utilization of Internet resources.

26

References

- IEEE Internet Computing Jan/Feb 2002
  Peer to Peer Networks in Action by Li Gong
- Java P2P Unleashed With JXTA, Web Services, XML,Jini, Java Spaces and J2EE (Sams Publishing)
- Peer-to-Peer Harnessing the Power of Disruptive Technologies, edited by Andy Oram.
- http://www.jabber.org/about/techover.php

27

References

- http://www.openp2p.com/pub/a/p2p/2000/11/24/shirky1-whatisp2p.html?page=1
- http://www.gca.org/papers/xmleurope2001/papers/html/s08-1.html
- http://wiki.cs.uiuc.edu/cs427/P2P+Architecture
- http://netserver.cerc.wvu.edu/classes/cs491h_summer2_2001/hunt_Intel_Security/ptptl.doc

28

Questions ?

29