



Towards Privacy-Enhanced Transactions for Virtual Organisations

Erica Y. Yang¹, Jie Xu¹, and Keith H. Bennett²

¹School of Computing, University of Leeds

²Dept. of Computer Science, University of Durham

All-hands Nottingham, September, 2004

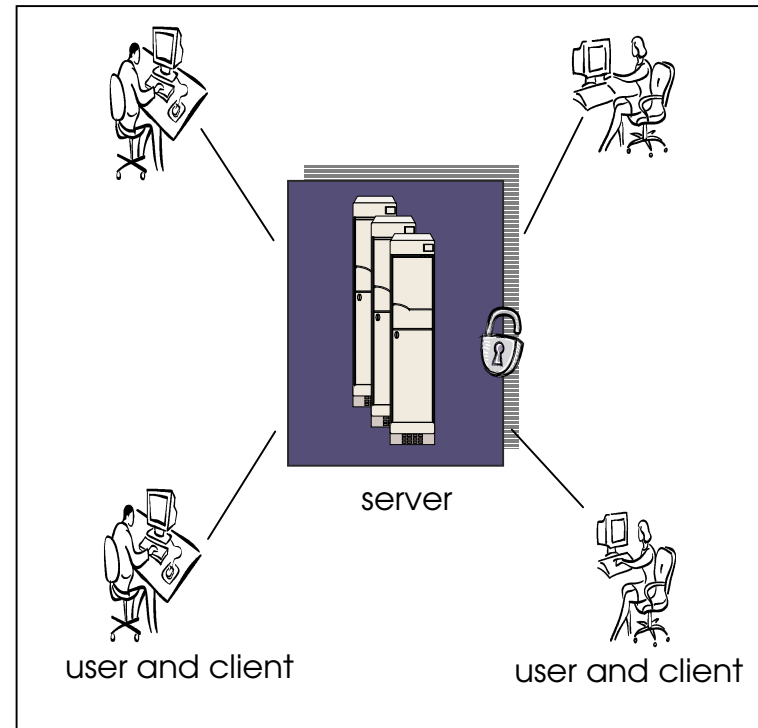
e-Demand Project:
<http://www.comp.leeds.ac.uk/edemand/>

Outline

- A typical security solution in traditional distributed systems
- Privacy protection in VO's – the new challenge
- Limitations of existing approaches
- Computing with encrypted data
- Privately Retrieving Information (**PIR**):
 - The single-server approach
- Preliminary results
- Technical implications
- Conclusions and future work

Security Issues in Andrew

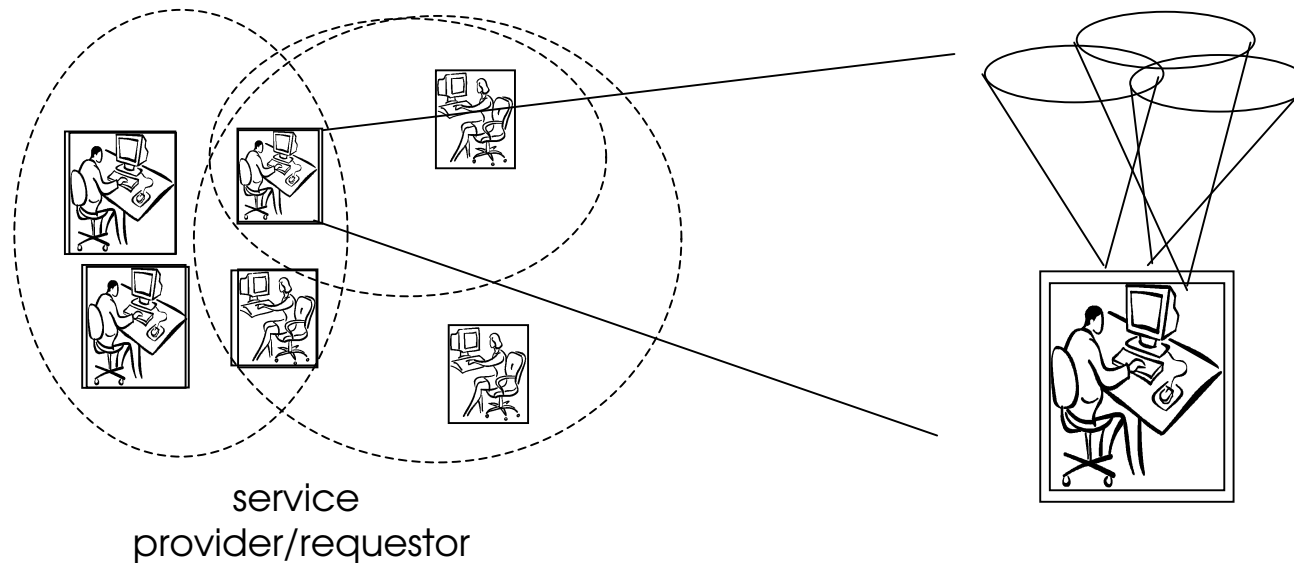
- **Andrew** is a distributed computing environment, built, operated and maintained by CMU.
- It is based on a centralised security paradigm
 - Isolated and physically guarded servers
 - At an abstract level (since it may be composed of many untrusted components), there is a single *trusted* authority (i.e. the server).
 - Users all have a long term trust relationship with the server.



→ What about Virtual Organisations (VO's)?

Virtual Organisations

- What are VOs?
 - “coordinated resource sharing and problem solving ...”
 - “dynamic boundaries and memberships, across multi-institutional domains ...”
- The heart of VOs is a dynamic relationship of sharing resources
- Trust management is one of the real challenges here.

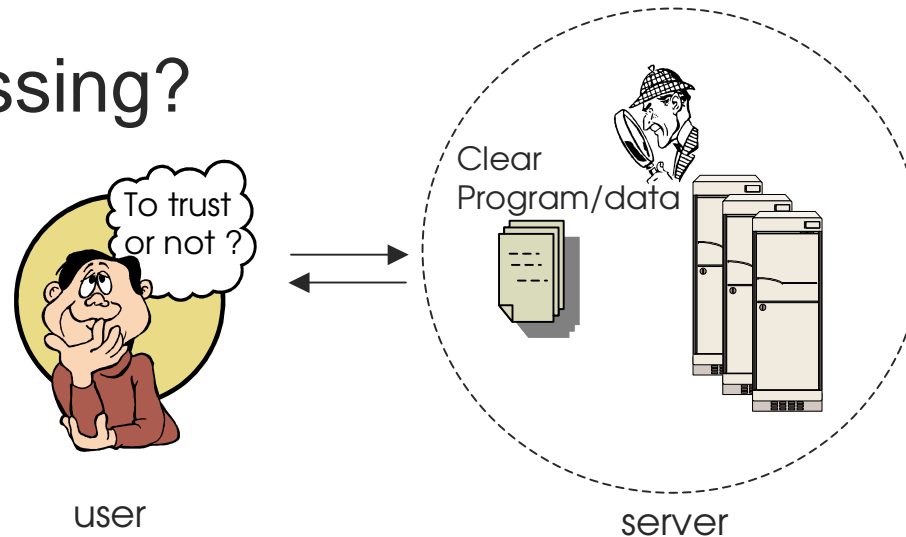


Trust Management and Implications

- What are the security-related implications of trust management?
 - Once a trust relationship is established between a user and a service provider, user's privacy could be in the hands of the service provider. **Andrew** is based on the same assumption implicitly. So, why is this a problem in VO's?
 - Because there is much less control over systems and services in VO's and so it is harder to achieve the degree of trust similar to that Andrew achieved
 - What about contracts and privacy policies?
- Indeed, what kind of privacy risks are we talking about?!

Privacy Risks of VO's

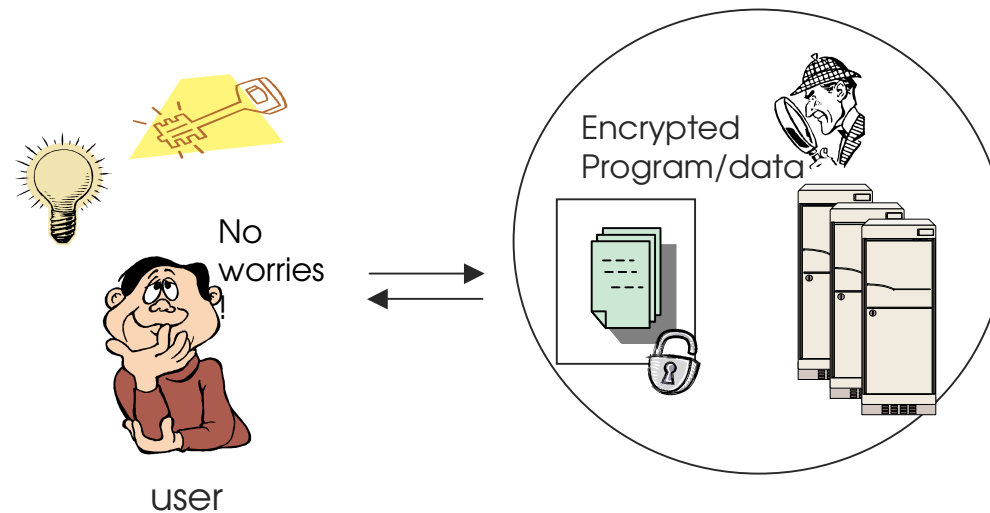
- Such risks are obvious in VO's as we have
 - Proprietary programs
 - Secret Information (e.g., private data, trade secrets)
- Sharing/utilising resources may compromise privacy
- What is missing?



A Missing Link

- What is missing is an enhanced security mechanism that can protect user's privacy during and after the computation on the server side. This should be assured even after the trust relationship has been established.
- The trust relationship in VOs is often transient and context-dependent.
 - The trustworthiness of service providers can *not* be naturally extended to the future.
 - No guarantee that members of a VO wouldn't (intentionally or unintentionally) leak confidential information to the other parties afterwards.

Encrypted Computations



- This is a proactive approach towards the trust problem without relying on the complete trust of servers
- Encrypted Computations
 - Computing with encrypted data
 - Executing obfuscated programs
- Only users know how to recover the intended results from those produced by the server using encrypted data

Single-server Private Information Retrieval (SPIR)

- SPIR is a data query technique for protecting the privacy (i.e., the intention) of a user based on two computationally intractable assumptions: the Discrete Logarithm Problem (DLP) and the factoring problem.
- A prototype implementation on a realistic database system is being under construction.
- Here is the principle.
 - User's intention: $f(i_a)$ e.g. 9^{1000}
 - Generating encrypted queries: $a = E(i_a, k)$ e.g. 9δ
 - Computing with encrypted queries: $b = F(a, i_b)$ e.g. $(9\delta)^{1000}$
 - Reconstructing the result: $f(i_a) = D(b, k, i_a)$ e.g. $(9\delta)^{1000} / \delta^{1000}$

Example (I) - Settings

- Suppose a database contains four values and a user wants to get the second data item (i.e., $i_a = 2$ and $x_{i_a} = 3$)
- Based on i_a , the user generates two prime numbers p and q , where $q|(p - 1)$. Let g be the generator of the subgroup G_q of Z_p^* . For example, $p = 23$, $q = 11$ and $g = 2$.
- The system keeps g, q private while sharing p with the database owner.

Position	1	2	3	4
Data	10	3	9	1

$$i_a = 2$$



$$i_1 = 0, i_2 = 1, i_3 = 0, i_4 = 0$$

Example (II) - Encrypting Queries and Computing with Encrypted Queries

- Generate query polynomials

Polynomial (mod 11)	$g_1(z) = z + i1$	$g_2(z) = 4z + i2$	$g_3(z) = 2z + i3$	$g_4(z) = 9z + i4$
z1=3	3	2	6	5
z2=5	5	10	10	1

- Prepare encrypted queries

(mod 23)	$2^{g_1(z)}$	$2^{g_2(z)}$	$2^{g_3(z)}$	$2^{g_4(z)}$
z1=3	2^3	2^2	2^6	2^5
z2=5	2^5	2^{10}	2^{10}	2^1

- Send encrypted queries to the database

(mod 23)	$2^{g_1(z)}$	$2^{g_2(z)}$	$2^{g_3(z)}$	$2^{g_4(z)}$
z1=3	8	4	18	9
z2=5	9	12	12	2

- Computing with encrypted queries on the database side

y1	$(8)^{10} (4)^3 (18)^9 (9)^1 = 13$
y2	$(9)^{10} (12)^3 (12)^9 (2)^1 = 18$

Example (III) – Result Reconstruction

- The user computes Lagrange co-efficiencies as follows.

$$c_2 = \frac{0 - z_1}{z_2 - z_1} \text{ mod } 11 = \frac{0 - 3}{5 - 3} \text{ mod } 11 = \frac{8}{2} \text{ mod } 11 = 8 \cdot 6 \text{ mod } 11 = 4 \text{ and}$$

$$c_1 = \frac{0 - z_2}{z_1 - z_2} \text{ mod } 11 = \frac{0 - 5}{3 - 5} \text{ mod } 11 = \frac{6}{9} \text{ mod } 11 = 6 \cdot 5 \text{ mod } 11 = 8$$

- The user then computes:

$$y = (y_1)^{c_1} \cdot (y_2)^{c_2} \text{ mod } 23 = 8$$

- The user finally keeps calculating for $R = 2^s \pmod{23}$ for $s = 1, 2, \dots$ until $y = R$. The corresponding s is the result x_{i_a} . In this case, $s = 3$, which is the value of the second data item. However, the database side has no way to figure out i_a , i.e. the user's real intention.

Technical Implications

- In this approach there is no trusted third party to enforce privacy protection. The user protects their own privacy (using encryption) while exploiting the resources of the service provider (e.g. retrieving the information from a remote database server).
- SPIR is based on computational intractability assumptions where a user can set the parameters on a session basis.
- To obtain a controlled level of privacy protection, users may decide the computational price they wish to pay. Therefore, the method is measurable and adjustable by users.

Other Considerations

- The need for a high degree of security protection depends very much on applications
- It is at the price of additional (computational) resource consumption
- It should be viewed as an extra level of security besides authentication, authorisation, and accountability (AAA)

Conclusions and Future Work

- Privacy protection is currently a weak link in the facilitation of Virtual Organisations.
- To fully understand the strengths and intrinsic limitations of our approach, we believe that extensive experimentations, both at an intranet and Internet scale, are important. We have obtained some important results
- The general applicability and feasibility of these technologies to a wider range of applications are still under investigation.
- (see our demo in a setting of replicated servers)