Panel Session
What are the key challenges in distributed security?

Steve Barker[1], David Chadwick[2], Jason Crampton[3],
Emil Lupu[4], Bhavani Thuraisingham[5]

[1]Dept. of Computer Science, King's College London, UK
[2]Computing Laboratory, University of Kent, UK
[3]ISG, Royal Holloway, University of London, UK
[4]Dept. of Computing, Imperial College London, UK
[5]University of Texas at Dallas, USA
{steve.barker@kcl.ac.uk, d.w.chadwick@kent.ac.uk,
jason.crampton@rhul.ac.uk, e.c.lupu@imperial.ac.uk,
bhavani.thuraisingham@utdallas.edu}

Abstract. The principal motivation for organizing a panel session at DBSEC'08 was to invite a number of distinguished researchers in data security to present their thoughts and to stimulate conference debate on a question of major importance: what are the key future challenges in distributed data security? The thoughts of the panellists on this issue are summarized in this article.

Steve Barker, the session moderator, opened the discussion by commenting that the term "distributed data security" describes a very wide-ranging space of issues that are often quite loosely related. For example, in terms of technologies, "distributed security" is applicable at the levels of the minute (e.g., hand-held devices) and the massive (e.g., the Internet). Barker noted that although there are common challenges (e.g., dealing with incomplete, contradictory, non-contemporary, and unreliable distributed sources) specific distributed systems present specific challenges. Barker also noted that the term "security" in "distributed data security" is also very general and covers privacy and integrity issues that present particular challenges in the distributed context. Barker concluded by observing that the contributions of the panellists revealed that the key challenges in distributed security remain many and varied.

David Chadwick argued that security will be increasingly policy-based with common policies being distributed to many sites so that a consistent approach to security can be developed throughout the system. Chadwick predicted that there would be significant advances in user-friendly tools for creating security policies and that these will be based on natural language so that humans will be able to understand clearly the policies they create. Many systems will have multiple stakeholders, each of whom will want to express their own security policies for (some of) the data in the system. Consequently there will be conflicting policies from the different stakeholders, which will require automated mechanisms for the resolution of policy conflicts. Chadwick suggested that federated identity management will increase in prominence, with single sign-on and attribute-based authorization, with the attributes coming from a variety of authorities.

Trusted platform modules will be utilized to increase trust between the federated systems. Users will become more aware of protecting their privacy as losses from identity theft increase. National ID-based schemes will be increasingly rolled out throughout Europe and will tend to be used for valuable transactions. Biometrics will be used more frequently for authentication. Biometric databases of entire populations will become more prevalent and will lead to increased fears of privacy leaks. (We already have the biometrics of 4 million people on the UK police database). Furthermore, networks will be patrolled by governments, police and the security services, and all traffic on the Internet will be routinely analysed. (Either legislation will be introduced to enforce ISPs to record all traffic, or it will be done surreptitiously at key gateways.) Chadwick noted that these developments will increase users concerns about privacy, making them turn to OpenID or similar systems, in which the users choose their own globally unique pseudonyms. There will be advances in anonymized data access for medical and other applications that require access to large distributed data stores of personal information, and intelligent history-based protection mechanisms will stop users from trawling and aggregating output in order to flout privacy rules.

Jason Crampton observed that access control models for closed, centralized environments assume the existence of components that are responsible for authenticating users, for intercepting requests and enforcing authorization decisions, and for deciding whether a request is authorized or not. Moreover, Crampton noted that, in the centralized case, it is assumed that mutual trust relationships exist between these components and that they share a common "vocabulary" for authentication and authorization.

Crampton expressed the view that implementing access control in open distributed environments can be very challenging because the assumptions that hold in the centralized case do not necessarily apply to decentralized systems. For example, prior trust relationships may not necessarily exist between components; indeed, they may not even be aware of each other's existence. Crampton suggested that five challenges emerge:

1. To be able to *map* a user in one domain to one or more principals defined in the authorization policy of another domain without any prior agreement between the domains.
2. To be able to *identify* all of the user attributes that are required to make an authorization decision.
3. To be able to *collect* all of the statements about user attributes that are required to make an authorization decision.
4. To develop a language to *encode* statements about user attributes in a common format with a universal semantics.
5. To be able to *verify* the authenticity of statements binding user identifiers to user attributes.

Emil Lupu suggested that the trend towards ”pervasive systems” leads us to envisage a world that includes mobile devices such as phones and PDAs, body area sensor-networks (e.g., for health monitoring), autonomous vehicles and instrumented environments such as smart-homes, autonomous buildings and watchful urban environments. Lupu noted that, in such environments, data is continuously acquired, aggregated and proactively exchanged amongst devices and amongst infrastructure services. Beyond access control, data protection requires privacy, dissemination and usage controls. Decisions regarding data protection, retention and disclosure need to be made in the presence of uncertain and partial authentication information and are often context dependent. Data exchanges are subject to regulations derived from legislation, organizational procedures and data sharing agreements between organizations. Expressing these, deriving operational policies, and deploying those policies to enforcement mechanisms close to the data remains a significant challenge. Policy analysis algorithms to detect and resolve conflicts between policies are also necessary. Frameworks in which data can be protected beyond the originator's domain need to cater for a variety of protection requirements and threat models. On smaller scale devices this needs to be achieved with limited computational resources. Yet the same techniques that are used for data protection may be abused to ensure its survival and proliferation.

Bhavani Thuraisingham observed that many technologies are being developed for distributed information management and that security and privacy issues have to be investigated in relation to these emerging technologies. Thuraisingham suggested that one of the main challenges in distributed information management is to support social networking algorithms and, for this, work on the integration of the information in disparate and diverse data sources is needed. In addition, the knowledge that is extracted from these information sources has to be integrated so that the manager(s) of them can make effective decisions. Today we see an explosion of social networks such as My Space and Face Book. Ensuring the security of access and privacy of individuals for such networks are critical issues. Thuraisingham reported that research at the University of Texas at Dallas (UTD) is focusing on developing novel and secure semantic web technologies for effective knowledge management and social networking. (Sponsors of this work include the Air Force Office of Scientific Research, the Intelligence Advanced Research Projects Activity, the National Science Foundation, the National Geospatial Intelligence Agency and Raytheon Corporation.) More specifically, a secure framework, based on the service oriented architecture paradigm, is being developed at UTD and is based on a three-level model that includes: The RDF Graph Manager, The Ontology Heuristics Reasoner and the Entity Extractor. Thuraisingham explained that novel dependable and secure semantic web technologies are being employed to realize this framework of connected layers. For example, the ontology-based heuristics reasoner will rely on the RDF graph manager to provide efficient storage and retrieval of RDF graphs. The entity extractor will depend on both the RDF graph manager and ontology-based heuristics reasoner to structure and reason about the graphs so that the entity extractor component can effectively carry out its task. All of the layers combined will provide the infrastructure support for distributed algorithms for social network analysis and knowledge management. Thuraisingham stated that one of the main focus areas for this work is security and privacy so that secure and private social networks can be supported. Thuraisingham noted that although research in secure distributed systems and distributed databases systems began in the 1980s, there remain many aspects of information distribution for which specific solutions for secure distributed networks, middleware, databases, information sources and applications are still needed. Thuraisingham concluded by suggesting that secure semantic web technologies will form the glue to secure various aspects of future distributed systems.