Secure Online Scientific Visualization of Atmospheric Nucleation Processes

Huaguang Song*, Weichao Wang†, Jinzhu Gao*, Aidong Lu†, and Lican Huang‡
*School of Engineering and Computer Science
University of the Pacific, Stockton, CA 95211
Email: h_song, jgao@pacific.edu
†Department of SIS, UNC Charlotte, Charlotte NC 28223
Email: weichaowang, alu@uncc.edu
‡Institute of Networking and Distributed Computing
Zhejiang Sci-Tech University, Hangzhou, P.R. China 310018
Email: licanhuang@zstu.edu.cn

***Abstract*—With the fast increases in the size of the scientific data, the visualization technique has been widely adopted to transform the information into an easy-to-understand representation. Since the security clearance and access rights of the end users may vary greatly in a scientific visualization system, the security mechanisms must be properly designed and deployed. In this paper, we present a key management and update approach for online visualization of atmospheric nucleation. The users are divided into multiple groups and the personal secrets are determined by combining the user identities and polynomials. The personal secrets support both user authentication and visualization result encryption. We also describe the stateless key update mechanism. The proposed approach has been integrated with our visualization system and tested with real scientific data.**

***Keywords*-Polynomial Based Personal Secret; Stateless Key Update; System Integration; Online Scientific Visualization.**

I. Introduction

Nucleation phenomena play a pivotal role in many atmospheric and technological processes. Understanding how particles or phases nucleate and grow in a multi-component mixture has important practical implications from climate to microemulsions, gas separations, and nano-materials. However, nucleation remains the least understood of steps influencing the concentration of aerosols, such as cloud condensation nuclei, in the atmosphere. Due to the lack of effective simulation approaches, the scientists have to depend on the large amount of sensing results as the analysis sources. Since the volume of the data can be really huge [1], the researchers have adopted the mechanisms such as scientific visualization to transform the raw information and analysis results into understandable figures or graphs.

While the storage and computation overhead of the data analysis for atmospheric nucleation usually restricts the processing procedures to a centralized data center, the end users of the raw data and analysis results can come from all over the world. Therefore, an approach to online visualization is highly demanded. In addition to the requirements on the performance and processing efficiency, the confidentiality and authenticity of the transmitted data must also be properly enforced in several application scenarios. For example, let us assume that the Department of Energy (DoE) at USA is collaborating with the researchers from another country A and is willing to grant them temporary access rights of data at a certain accuracy level. Both authentication of the readers and encryption of the data must be conducted during the access.

This application scenario can be abstracted as the model illustrated in Figure 1. The users are divided into multiple groups and each group may have different access rights to the visualization results. The data center needs to verify the identity of the user. The visualization results will be encrypted with a secret known only to the data center and the user. At the same time, the data decryption overhead at the end user and the key management overhead at the data center should be very limited.

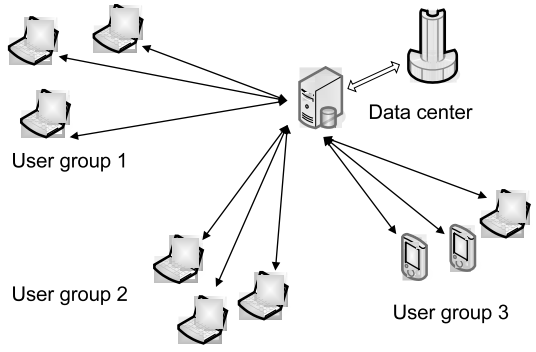

Figure 1. Data access model of the investigated application.

For this application, the computation overhead of the traditional asymmetric encryption schemes could be too heavy for some end users. At the same time, if the data center has to remember the encryption key for every user, the storage overhead could be very heavy as well. Therefore, in this paper we propose a polynomial based key derivation protocol to manage data access of different groups of users. Here the data center will generate a different polynomial for

each user group and the personal secret can be determined by applying the user identity to the polynomial. The robustness of the approach can be controlled by adjusting the degree of the polynomial. We also design a stateless key update mechanism to support secret refreshment and avoid the same keys to be used for too long. The analysis shows that the proposed approach will cause very limited overhead at the end users.

We have implemented the proposed key management mechanism and integrated it with our online visualization system for the atmospheric nucleation. The online system has a web-based interface to facilitate effective remote visualization and analysis of atmospheric nucleation processes. Intelligent control mechanisms are used for remote file access, interactive visualization, as well as simultaneous data access. By utilizing the key management mechanism, we are able to provide a secure environment for nucleation scientists to access our online system, which is essential but typically missing in similar systems.

The remainder of the paper is organized as follows. In Section II we introduce the related work. In Section III we present the details of the key management and update mechanisms. We also analyze the overhead of the proposed approach. In Section IV we introduce the integrated system. Finally, Section V concludes the paper.

II. Related Work

Over the years, advanced middlewares and frameworks have been developed to facilitate collaborative data analysis and visualization. For example, Brodlie et al. [2] assessed a selection of visualization systems and frameworks for their use in a collaborative environment. Grimstead *et al.* [3] presented a collaborative grid enabled visualization environment that supports automated resource discovery. Many Eyes [4], a public website for uploading data and creating interactive visualization, supports collaborative visualization at a large scale.

With the increasing challenges of data analysis and visualization, collaborative problem-solving has started to draw more attention of visualization researchers. For example, Park et al. [5] explored collaboration issues for a CAVE-based virtual reality environment. Waldner et al. [6] discussed design considerations for employing multiple-view visualizations in collaborative multi-display environments. Bresciani and Eppler [7] analyzed the impact of visualization on knowledge sharing in situated work groups and showed that interactive visualization could bring positive and productive for group work.

Polynomial interpolation was first used to implement threshold secret sharing [8]. It allows a dealer to distribute a secret s to n players and at least $t < n$ players are required to recover the information. Staddon *et al* [9] proposed a self-healing key distribution mechanism with revocation capability based on the secret sharing techniques. Liu *et al* [10] proposed an efficient self healing group key distribution scheme with revocation capability based on the result. In [11], the authors proposed a key distribution and update mechanism for group based information access. These methods provide a solid foundation for our approach.

III. Data Access and Key Update for Secure Online Visualization

A. Assumptions and Notations

We adopt a simplified model to describe the user groups of the visualization system. We assume that the end users can be divided into multiple groups and different groups may have different access rights to the data. The data center will verify the identity of the user before sending the encrypted data to her/him. The secret keys of the users in the same group need to be updated when the group member changes. At the same time, it should be fairly easy for the data center to remove all the secret keys associated with a user group. During the lifetime of a user group, we assume that at most l users will leave the group. These users should not get further data access rights after they leave the group.

We assume that the attackers can eavesdrop on and record the packets that are transmitted over the network. They can also conduct active attacks by inserting, modifying, or discarding packets. We assume that the attackers do not have the computation resources to directly break the encryption keys. Some of the users in the system are curious and they try to read the visualization results that are sent to other users. Mechanisms must be designed to defend against such attacks.

We assume that every user can be uniquely identified by the name u, where $u \in \{1 \cdots n\}$ and n is the total number of users. The users are divided into d different groups, which are represented by G_1 to G_d respectively. All operations described in the paper will take place in a finite field F_q, where q is a prime number with a large enough value.

We use $E_k(msg)$ and $D_k(msg)$ to represent the encryption and decryption of the message msg with a symmetric key k respectively. We use $h(x)$ to represent a t-degree polynomial in $F_q[x]$ and $h(u)$ is the value of the function at point u. Similarly, we use $F(x)$ to represent the $l+t$ degree polynomials, the functionality of which will be described later. We assume that a user group will conduct at most m rounds of key updates.

Table I summarizes the notations used in this paper. We assume that u is a member of G_v. The key encryption keys and the refreshment operations will be discussed in the next subsection.

B. Secure Visualization Access

During the network initiation procedure, every user will receive a set of secret keys from the data center through a secure channel. For example, DoE can provide the secret keys to the scientists from country A when the collaboration

Table I
NOTATIONS OF THE PAPER.

u	user u's identity
DC	the data center
G_v	the user group v
m	at most m rounds key updates for each user group
l	at most l users will leave a group
$h_{v,j}(x)$	t-degree polynomial determining personal secrets of user group G_v in round j
$F_{v,j}(x)$	$(l+t)$-degree polynomial used for personal secret updates of user group G_v in round j

starts. These keys can be divided into two groups: traffic encryption keys (TEK) to protect visualization results, and key encryption keys (KEK) to support secret refreshment. Below we use a user u in group G_v as an example to illustrate the data access operations.

We use the t-degree polynomials $h(x)$ to determine the personal key shares and protect the visualization results. The data center will use a polynomial $h_{v,j}(x)$ to determine the personal keys of the users in group G_v. Here the first and second indexes represent the user group and the number of key updates that have been conducted for the group. Therefore, we have $v \in \{1 \cdots d\}$ and $j \in \{1 \cdots m\}$. For example, $h_{v,1}(x)$ is the polynomial to determine the personal key shares of the members in G_v when they first join the system. A user u in G_v will get its personal key share $h_{v,1}(u)$. When it wants to send a data access request to the data center, it will send out:

$$(u, Data\ Center, E_{h_{v,1}(u)}(request)).$$

Since the data center knows the polynomial $h_{v,1}(x)$, it can calculate the personal key share $h_{v,1}(u)$ by applying u to the polynomial and recover the information. Similarly, the data center can encrypt the visualization results with the personal key of u and send them back.

Two advantages have been brought by the personal key shares determined by the polynomials. First, for two different users u and w in G_v, they will have different personal keys $h_{v,j}(u)$ and $h_{v,j}(w)$ to encrypt the visualization results. Therefore, information isolation has been achieved, and only the data center and the targeted receiver can recover the information. Second, it becomes more difficult for an attacker to impersonate another user in the same group unless it can collect $t+1$ personal keys and reconstruct the polynomial $h(x)$. At the same time, the user will not have any information about the secret keys of other groups.

C. Stateless Key Updates

When a group change happens, the corresponding keys must be updated to enforce information secrecy. In this part, we present the approach to stateless key updates for secure communication between the users and the data center. We first introduce the pre-distributed information that is used in secret update and recovery. Then we describe the key update operations for the members of a group.

C.1 Pre-distribution

To support stateless key refreshment, the data center will distribute some information that is essential to the secret recovery operations to a user during the system initiation procedure. We assume that the users are divided into d groups, and each group will conduct at most m rounds of secret updates. Within the same group, at most t users will collude together to impersonate another member. At the same time, at most l users who were members of the group G_v will leave the group, and their user names will be put in the revocation set $R_{v,j}$. Here the first index represents the user group, and the second index denotes the key update round number. The numbers t and l will jointly determine the degrees of the polynomials that are used during key refreshment.

The data center will select m polynomials with the degree $(l+t)$ from $F_q[x]$ for each group to serve as the 'masking functions' for personal key update operations. Every function is denoted as $F_{v,j}(x)$, where $v = 1 \cdots d$, and $j = 1 \cdots m$. Here v represents the user group, and j represents the round number of key update in group G_v. Every user u in group G_v will receive the values $F_{v,j}(u)$, $j = 1 \cdots m$. The pre-distributed key encryption information is also summarized in Table I.

C.2 Key update operations

For the group G_v, when the secret keys of the users need to be updated, the data center can broadcast a message and allow the users to recover their new personal secrets. This procedure can be achieved through true broadcast and does not need the users to contact the data center individually to get their new personal keys. Any users that have left the group will not be able to recover the new personal secrets even when they combine their pre-distributed information. At the same time, because of the stateless property, a user can recover the next round personal secret even if she/he misses the previous round key update message.

For the group G_v, given the set of revoked users $R_{v,j} = \{v_1, v_2, \cdots, v_p\}$, $p \leq l$, the data center will broadcast:

$$\begin{aligned}(\quad & DC,\ G_v,\ personal\ key\ update\ for\ G_v\ in\ round \\ & j,\ R_{v,j},\ \overline{P}_{v,j}(x) = g_{v,j}(x) \cdot h_{v,j}(x) + F_{v,j}(x), \\ & digital\ signature\ of\ DC\ )\end{aligned}$$

where $g_{v,j}(x)$ is determined by the names of the revoked users as $g_{v,j}(x) = (x - v_1)(x - v_2) \cdots (x - v_p)$.

Every user u in G_v that does not leave the group will try to recover the new personal key $h_{v,j}(u)$ from the received packet. It can calculate $\overline{P}_{v,j}(u)$ and $g_{v,j}(u)$ by applying its user name to the polynomials. Since u has received $F_{v,j}(u)$ during the system initiation procedure, it can calculate $h_{v,j}(u) = \frac{\overline{P}_{v,j}(u) - F_{v,j}(u)}{g_{v,j}(u)}$. For any user $y \in R_{v,j}$, since $g_{v,j}(y) = 0$, it cannot recover the new personal key.

When a new user joins the group, the data center needs to provide some information to her/him. We assume that user w

joins group G_v in round j. The data center will provide the keys to w through a secure communication channel between the two entities. The secrets will include the personal key share $h_{v,j}(w)$, and the values of the masking functions $F_{v,j'}(w)$ $(j' = j \cdots m)$.

Since the safety of the proposed mechanism heavily depends on the quality of the secrets and coefficients of the polynomials that are generated by pseudo-random number generators, below we discuss the generation of these parameters. In our application, the secret keys and polynomials can be generated off-line during the system initiation procedure. Under this condition, the generation procedure is not restricted by the computation overhead, and those strong yet complicated generators can be adopted [12], [13].

D. Analyses

In this subsection, we investigate the storage and computation overhead of the proposed mechanism at each user. Although the data center generates many polynomials in the proposed mechanism, the information that every user needs to store will take only a small space. Every user needs to store one personal key, and at most m values of the masking functions. The proposed mechanism will cause a limited amount of computation overhead at the users. To recover the new personal key, a user needs to evaluate a few polynomials and verify a digital signature of the data center to prevent the attackers from generating fake key update packets. Most of the operations, except for digital signature verification, can be accomplished efficiently [14]. Verifying digital signatures will not cause a large amount of computation overhead when elliptic curve based approaches are adopted [15].

The key update mechanism described in this paper has the stateless property: recovering the latest personal key shares does not depend on the knowledge of keys for previous sessions. This feature is especially important for the application scenarios in which the packet delivery cannot be guaranteed. If the personal keys have to be recovered sequentially, more and more users will not be able to decrypt the visualization results unless they initiate individual requests to get the latest secrets from the data center.

IV. System Implementation

To demonstrate the effectiveness of the proposed key management mechanism, we applied it to our online visualization system for the atmospheric nucleation. Our goal is to give remote users a secure collaborative workspace that promotes large-scale data exploration and knowledge discovery.

A particle-based nucleation simulation typically produces two kinds of data sets. The first type of data defines the nucleation free energy surface with a table of three columns representing number of water molecules, number of hexanol molecules, and the corresponding nucleation free energy

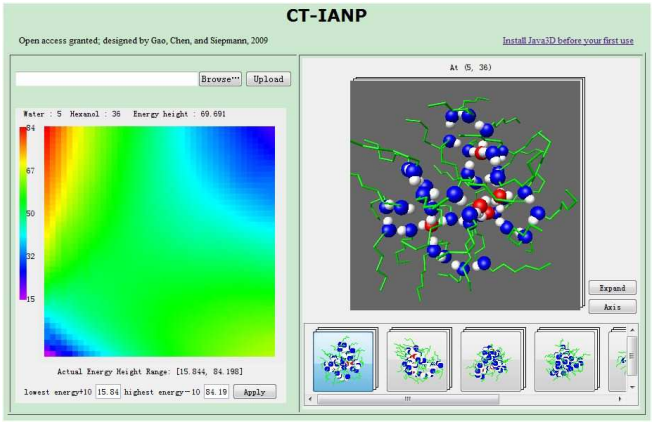

Figure 2. The graphical user interface of the online system.

respectively. The second type of data consists of the coordinates of molecules as well as atoms in aggregate structures. Figure 2 demonstrates the user interface design of the online system. The 2D image on the left shows the nucleation free energy contour map [16]. The color at each pixel on this map implies the magnitude of the nucleation free energy for the particular aggregate that this pixel refers to, which is controlled by a color scale bar. When the user moves the cursor over this interactive map, the top side bar would show the aggregate composition (i.e., the number of water and hexanol molecules) as specified by this point on the map and its nucleation free energy. On the right, molecular structures for all possible configurations for this aggregate are displayed. File folders demonstrate the clustering results for pattern recognition. The user could open a folder to view the configurations that have strong similarities in features such as molecular topology.

Following the Model-View-Controller design pattern, the online system consists of three modules: (1) Visualization module, which is in charge of displaying visualization results on a webpage which consists of html, JSP and applets. Two applets that could communicate with each other are implemented. One applet handles the information about the raw data, including the description of the molecule types in the binary mixture, the current free energy range and its color scale bar, boxes to change the free energy range, a 2D color map of the free energy landscape, the location of the cursor specified in terms of the number of water and hexanol molecules, and the free energy for this aggregate size. The other applet supports the 3D visualization of aggregate structures containing a specific number of water and hexanol molecules. A user is able to perform operations such as Zoom in/out, translation, or rotation in the display panel. (2) Data module, which parses the data file. As a aggregate structure data file may contain more than thousands of configurations, the response to a user query for a particular aggregate in such a big file will definitely

be very slow. To solve this problem, during preprocessing stage, we partitioned the big file into smaller files, indexed by the number of hexanol molecules and the number of water molecules. (3) Control module, which is responsible for web-based scheduling, data forwarding as well as error handling. The visualization module sends requests to the control module for the display content while the control module sends requests to the data module for the data that is required for satisfying the display request.

To support the secure system design, an administrative interface is added for user and group management. Only the system administrator has the right to create a new group and grant a group with suitable access rights and a unique polynomial. Under the security enhanced system design, before accessing the functionalities supported by the system, each user is required to register and make a request to join a group. The system administrator will review each request and decide whether to accept or deny. A unique user identity is automatically generated and assigned to the user if his/her request is accepted.

Efficiency and scalability have been the two major concerns in our system design. Limited amount of the storage and computation cost makes the proposed key management mechanism an ideal security solution for our online system to support data analysis and visualization needs of nucleation researchers worldwide. The solution should also be applicable to other similar online tools.

V. Conclusion

Secure online visualization is an important component of many scientific applications. In this paper, we focus on key distribution and update for secure information access in atmospheric nucleation visualization. The proposed mechanism adopts polynomials to support the distribution of personal key shares and employs stateless secret update to achieve efficient key refreshment. It becomes more difficult for an attacker to impersonate another entity in the network. The proposed mechanism introduces only a small amount of storage and computation overhead to the users. In the future, we plan to integrate the proposed mechanism with other visualization systems to evaluate its performance in more scientific applications.

References

[1] J. Smith, M. Dunn, T. VanReken, K. Iida, M. Stolzenburg, P. McMurry, and L. Huey, "Chemical composition of atmospheric nanoparticles formed from nucleation in tecamac, mexico: Evidence for an important role for organic species in nanoparticle growth," *Geophys. Res. Lett.*, 2008.

[2] K. Brodlie, D. Duce, J. Gallop, J. Walton, and J. Wood, "Distributed and collaborative visualization," *Computer Graphics Forum*, vol. 23, no. 2, pp. 223–251, 2004.

[3] I. J. Grimstead, N. J. Avis, and D. W. Walker, "Automatic distribution of rendering workloads in a grid enabled collaborative visualization environment," in *SC '04: Proceedings of the 2004 ACM/IEEE conference on Supercomputing*. Washington, DC, USA: IEEE Computer Society, 2004, p. 1.

[4] F. B. Viégas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon, "Many Eyes: A site for visualization at internet scale," in *Proceedings of IEEE InfoVis 2007*, 2007.

[5] K. S. Park, A. Kapoor, and J. Leigh., "Lessons learned from employing multiple perspectives in a collaborative virtual environment for visualizing scientific data," *Proceeding of Collaborative Virtual Environments*, pp. 73–82, 2000.

[6] M. Waldner, A. Lex, M. Streit, and D. Schmalstieg., "Design considerations for collaborative information workspaces in multi-display environments," *Proc. of Workshop on Collaborative Visualization on Interactive Surfaces*, 2009.

[7] S. Bresciani and M. J. Eppler, "The benefits of synchronous collaborative information visualization: Evidence from an experimental evaluation," *Information Transaction on visualization and computer graphics*, pp. 1073–1080, 2009.

[8] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, pp. 612–613, 1979.

[9] J. Staddon, S.Miner, M. Franklin, D. Balfanz, M. Malkin, and D. Dean, "Self-healing key distribution with revocation," in *Proc. of IEEE Symposium on Security and Privacy*, 2002.

[10] D. Liu, P. Ning, and K. Sun, "Efficient self-healing group key distribution with revocation capability," in *Proc. of ACM conference on Computer and communications security*, 2003, pp. 231–240.

[11] W. Wang and T. Stransky, "Stateless key distribution for secure intra and inter-group multicast in mobile wireless networks," *Elsevier Computer Networks*, vol. 51, no. 15, pp. 4303–4321, 2007.

[12] F. Panneton and P. Lecuyer, "On the xorshift random number generators," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 15, no. 4, pp. 346–361, 2005.

[13] F. Panneton, P. Lecuyer, and M. Matsumoto, "Improved longperiod generators based on linear recurrences modulo 2," *ACM Transactions on Mathematical Software*, vol. 32, no. 1, pp. 1–16, 2006.

[14] P. Ni and Z. Li, "Energy cost analysis of ipsec on handheld devices," *Microprocessors and Microsystems, special issue on Secure Computing Platform*, vol. 28, no. 10, pp. 585–594, 2004.

[15] H. Wang, B. Sheng, and Q. Li, "Elliptic curve cryptography based access control in sensor networks," *International Journal of Sensor Networks*, vol. 1, no. 3/4, pp. 127–137, 2006.

[16] R. Nellas, S. Keasler, and B. Chen, "Molecular content and structure of aqueous organic nanodroplets from the vapor-liquid nucleation study of the water/n-nonane/1-alcohol series," *J. Phys. Chem. A 112*, pp. 2930–2939, 2008.