ISSN 2186-7437

NII Shonan Meeting Report

No. 2012-8

Grid and Cloud Security: A Confluence

Barton P. Miller
Elisa Heymann
Yoshio Tanaka

October 15–18, 2012

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Grid and Cloud Security: A Confluence

Dates: October 15 - 18, 2012

Organizers

Prof. Barton Miller, University of Wisconsin, USA

Prof. Elisa Heymann, The Autonomous University of Barcelona, Spain

Dr. Yoshio Tanaka, National Institute of Advanced Industrial Science and Technology, Japan

Overview of the meeting

The security of Grid and Cloud computing environments is critical to today's cyber-infrastructure. The goal of this seminar is to bring together a diverse community of researchers, practitioners, and developers to leverage knowledge that spans the areas of Grid and Cloud security, industry and government and academia, theoretical and practical interests, and the scientific and business communities. This Shonan meeting will make use of the unique opportunity of its week-long format to bridge these areas and establish long-lasting research collaborations between the various communities. The seminar will comprise both representative background presentations to set the context for discussions, working sessions to develop joint research agendas, and sessions that focus on joint problem-solving of a target issue selected during the week.

1. Participants

Title	Name	Affiliation
Dr.	Akihito Nakamura	AIST (JP)
Dr.	Akira Otsuka	AIST (JP)
Prof.	Atsuhiro Goto	Institute of Information Security (IISEC) (JP)
Prof.	Barton Miller	University of Wisconsin (US)
Dr.	David Presotto	Google (US)
Associate Professor	Eisaku Sakane	National Institute of Informatics (JP)
Prof.	Elisa Heymann	Universidad Autónoma de Barcelona (ES)
Prof.	Gene Tsudik	University of California, Irvine (US)
Mr.	Hiroki Hada	Institute of Information Security (IISEC) (JP)
Mr.	James Kupsch	University of Wisconsin (US)
Dr.	John White	Helsinki Institute of Physics (FI)
Prof.	Kento Aida	National Institute of Informatics (JP)
Dr.	Linda Cornwall	Rutherford Appleton Lab (UK)
Mr.	Loren Kohnfelder	Google (US)
Ms.	Rika Hayashi	Institute of Information Security (IISEC) (JP)
Dr.	Shiho Moriai	NICT (JP)
Assistant Professor	Shin'ichiro Takizawa	Tokyo Institute of Technology (JP)
Prof.	Shinji Shimojo	Osaka University (JP)
Dr.	Stefano Zatti	European Space Agency (IT)
Mr.	Takamichi Asou	Institute of Information Security (IISEC) (JP)
Dr.	Yoshio Tanaka	AIST (JP)

2. Schedule

Monday, October 15th

Monday morning, 9:00am – 10:30am: Opening and Vulnerability Assessment

(1) Welcome from Prof. Zhenjiang Hu, Academic Committee Chair of NII Shonan Meeting

(2) Brief discussion of meeting organization and format; assignment of note takers.

(3) James Kupsch: "Experiences with Middleware Vulnerability Assessment"

Monday morning, 11am – 12:30pm: Vulnerability Assessment in the MIST Project

(1) Barton Miller: "Automated Tools for Threat Model Extraction"

(2) Elisa Heymann: "Automating Risk Analysis of Software Design Models"

Monday afternoon, 2:00pm – 3:30pm: Authentication and Crypto

(1) John White, "Securing the Grid, EGEE to EMI and beyond"

(2) Shiho Moriai, "Lightweight Cryptography for the Cloud: Exploiting the Power of a Bit Slice Implementation"

Monday afternoon, 4:00pm – 5:30pm: Protocols

(1) Eisaku Sakane, "A Study of On-line Interactions between Public Key Infrastructure Components"

(2) Akira Otsuka, "e-Voting in the Cloud: Information-theoretic Security through an Honest Majority"

Tuesday, October 16th

Tuesday morning, 9am – 10:30am: Privacy

(1) Gene Tsudik: "Hummingbird: Privacy at the Time of Twitter"

(2) Loren Kohnfelder, "Cloud computing privacy perspectives"

Tuesday morning, 11am – 12:30pm: Cloud Security

(1) Atsuhiro Goto, "Inter-Cloud Computing for Secure Social Infrastructure: What's Next?"

(2) David Presotto, "Why should we trust cloud computing?"

Tuesday afternoon, 1:30pm – 3:00pm: Vulnerabilities and Intrusions

(1) Akihito Nakamura, "Unified Vulnerability Management SaaS Based on Open Standards"

(2) Takamichi Asou: "An Analytical Study for Sensor Service Provider to Keep their User's Security"

Tuesday afternoon, 3:30pm – 5:00pm: Policies and Experience

(1) Linda Cornwall, "Grid security policies, procedures, and activities: defining responsibilities and making things happen"

(2) Yoshio Tanaka, "Migrating from Grid to Cloud: GEO Grid's experience from Security Perspective"

Wednesday, October 17th

Wednesday morning, 9am – 10:30am: New Environments

(1) Stefano Zatti: "Clouds in Space? The approach of the European Space Agency to reach the clouds"

(2) Shinji Shimoto, "Can the New Generation Networking Idea and its Testbed Help to Improve Security?"

Wednesday morning, 11am – 12:30pm: Mobile Security and Malware Analysis

(1) Rika Hayashi, "Android Security Improvement by Visualization of Application Behavior"

(2) Hiroki Hada, "Using a Database in the Cloud for the Static Analysis of Malware"

Thursday, October 18th

Thursday morning, 9am – 10:30am: Infrastructures

(1) Kento Aida: "Authentication System for High Performance Computing Infrastructure in Japan"

(2) Shinichiro Takizawa, "VM Hosting for High Performance Computing Infrastructure in Japan"

Thursday morning, 11am – noon: Closing Discussions

(1) Discussion of future directions and report outlining.

3. Session details; abstracts and Q&A

Monday, October 15th

Monday morning, 9:00am – 10:30am: Opening and Vulnerability Assessment

(1) James Kupsch: "Experiences with Middleware Vulnerability Assessment"

abstract: Vulnerability assessment is a crucial aspect of assuring the security of software. For last seven years we have performed vulnerability assessments of grid middleware and other pieces of software. This talk presents the methodology we developed for performing assessments, and interesting facts we learned from our experience.

Q&A

Was Globus assessed?

Not directly, as the folks at Globus were not interested. But indirectly some components were assessed when we assessed Condor, though not Globus core.

What does it mean an unsafe file open?

Example: O_CREATE flag without O_EXCLUSIVE vulnerable to race conditions. Files in /tmp. Also problems depending on the desired semantic.

Comment on randomization of address spaces.

(After Jim mentioned some vulnerabilities cannot be exploited because of a bug preventing that!)

Is defensive programming the same as secure programming?

Basically, it is the same thing. Defense programming is being careful. Secure programming is defensive programming plus extras, such as preventing race conditions ...

Comments on programming languages?

Certain languages have different kinds of vulnerabilities. For example, using random packages with Java. Howver some vulnerabilities are common to all languages. Problems can arise from using scripting languages as glue. Perl is quite problematic for security.

How do you communicate with the software development team?

Sometimes face to face. Lots of e-mail exchange. Usually documentation is of limited use as it is often out of date or incomplete. The initial build of a piece of software is difficult, but after that's done, there's not a lot of need to talk with the developers. We always do blue-team activities (as opposed to red team).

Discussion about disclosing vulnerabilities.

Fortify is widely deployed. How do you do the analysis?

We are comparing the automated tools against a manual assessment.

Do the assessors need malpractice insurance?

That's an interesting question but not one that we have talked about. Many technical practitioners have commercial liability insurance, so that might be reasonable here.

Monday morning, 11am – 12:30pm: Vulnerability Assessment in the MIST Project

(1) Barton Miller: "Automated Tools for Threat Model Extraction"

abstract: Visualizing a program's structure and security characteristics is the intrinsic part of in-depth software security assessment. Such an assessment is typically an analyst-driven task. The visualization for security analysis is usually labor-intensive, since analysts need to read documents and source code, synthesize trace data from multiple sources (e.g., system utilities like lsof or strace). To help address this problem, we propose SecSTAR, a tool that dynamically collects the key information from a system and automatically produces the necessary diagrams to support the first steps of widely-used security analysis methodologies, such as Microsoft Threat Modeling and UW/UAB First Principles Vulnerability Assessment (FPVA). SecSTAR uses an efficient dynamic binary instrumentation technique, self-propelled instrumentation, to collect trace data from production systems during runtime then automatically produces diagrams. Furthermore, SecSTAR allows analysts to interactively view and explore diagrams in a web browser. For example, analysts can navigate the diagrams through time and at different levels of detail.

We demonstrated the usefulness of using SecSTAR to produce FPVA-style diagrams for a widely used and complex distributed middleware system, the Condor high-throughput scheduling system. Compared with the original manual approach in FPVA, SecSTAR shortened the initial diagram construction time from months to hours and constructed a more accurate diagram visualizing the complete runtime structure of Condor.

Q&A

Stefano Zatti: What is the difference between a classical flowchart used in software development and flowcharts you produce?

Charts produced in FPVA are more course grained and labelled for security, providing a higher level view of component interaction. There are many different possible visualizations.

How do your diagrams relate to those from Microsoft's methodology?

Microsoft requires all developers to produce data flow diagram (DFD) in advance, during the design phase; the FPVA diagrams that we described are produced from the code after it is written.

In the Condor self-propelled instrumentation demo, why is one of the processes constantly changing color?

This job is starting other jobs, using privilege escalation/de-escalation. It is saying 'hack me' and a designer should put special effort into protecting it.

Yoshio - How many events do you support capturing?

Currently we support a lot but by no means all; however, it's trivial to add more. If system call isn't traced that you are interested it, it is very easy to add support with just a couple of lines of code. Clone (a kind of super fork) is one example of an unsupported event ... it's tricky because it has so many arguments.

Dave - That was my fault ... or rather they Linux folks didn't copy the Plan 9 implementation correctly.

Exactly how do you trap system calls, with ptrace?

Binary code injection up to the point of the transition into the kernel. The self-propelled trace implementation currently uses a simple form of binary code instrumentation that assumes there is no dynamic (self-modifying) code. If you wanted to relax that assumption, it isn't hard - just use a more complex form of DynInst [http://www.dyninst.org/], but there is a certain cost to do that.

When a vulnerability is exploited, can it be found by the technique?

Now, not at all. We are not looking for vulnerabilities at this stage, only looking at how processes operate and their control flow and privileges. It would be interesting to combine with tool that looks for static vulnerabilities, currently analyst finds vulnerabilities.

How do you choose test cases to run to achieve good coverage?

We try to work with the software team to come up with good representative set. E.g., we iterate so that if a process has been missed, we make sure to cover it with another tests case. No formal code coverage. We will also be investigating combining static and dynamic; combine the merged set to see what is missed. This will be tried next semester.

(2) Elisa Heymann: “Automating Risk Analysis of Software Design Models”

abstract: Fixing software security issues early in the development life-cycle of applications reduces its cost dramatically. Companies doing software development know this reality, and they have introduced risk assessment methodologies in their development processes. Unfortunately, these methodologies require engineers to have deep software security skills to carry out some of the most important steps of this process, and training them on security is expensive. In this scenario, we propose a new automated approach to analyze software designs to identify, rank and mitigate potential threats to the system. We designed a new data structure to detect threats in software designs called Identification Trees. We also defined a new way for describing countermeasures to threats using Mitigation Trees. Our automated approach relies on Identification Trees and Mitigation Trees to integrate a guided risk assessment process through the development life-cycle. It does not require developers to have any security training, and was integrated in the current Threat Modeling process of Microsoft.

Q&A

Stefano – For your cost rank of 1 to 10, how do establish these values?

In an ad hoc way and this is probably the weakest point at this stage, there isn’t good info available.

Bart: We welcome ideas from any in this area.

Within grid community it is notable that there is no strong agreement on what assets are more critical; it depends on the application. E.g. biomedical data needs strong protection, but physics data just needs strong integrity.

How do you determine the coverage of your threat analysis?

A lot of work qualifying existing threats based on manual assessment - a long list of complex vulnerabilities found. Tools are helping rather than taking human out the loop.

Facilities for qualifying new threats to be codified into the system.

End of Session Follow-Up Discussion

Gene - Has any of your focus been on assessing security software and protocols, since vulnerabilities can creep into any step.

Bart – You can find bugs at any level, e.g. at the hardware level attackers are now looking at ways of making CPUs incorrectly execute.

Chrome uses many different libraries (from open source community).

James Kupsch - Chrome has about 20 million lines of code, but it depends what of this you consider Google code.

Bart – It is necessary to define a boundary and assume things work beyond there, you cannot look at all code. E.g. A specialized community is looking at protocols.

Use between code and use of libraries. Most common attacks are not on crypto itself, but on usage of crypto. E.g. German navy not changing key for Enigma, using girlfriend's birthday as long-term key facilitate breaking.

Open source - people who specialize at looking for certain types of vulnerability will try and attack.

Better to have many eyes on code. Many eyes are not a substitute for a good assessment.

Dave - Google has no independent assessment methodology within the company.

Bart - One product people keep asking us to produce is students who can do assessments, there are a lack of skills in this area.

E.g. Does this software despite secure programming practice have any exploitable vulnerabilities? Does it adhere to standards?

Dave - A big problem is that often there is no clear specification of correct behavior.

Bart - First diagram we did was in Condor. Do analysis by looking at what are the most critical parts - look at value of it, find how much could be done if exploited. It's a matter of economics - how much do you want to pay?

Common question: If we can't afford analysts what tool should we use?

Dave - By the time analysis has completed - Google will have changed at least 5 times.

Crowdsource - people who want to look at software, difficult to decide who you can trust.

When people identify a vulnerability - some consider whether they will get more money by selling it to the company or to an intelligence agency.

Found a lot errors in Google that are not so pretty. Maybe get to crash software, but by good top level design ability of a specific part to affect another is restricted. This is due to good architecture (the canvas sandbox).

Errors found - generally 2/3 not worth cost to fix. If a tool is smart, it can find more vulnerabilities. Constant ongoing assessment of tools against skilled analyst.

Intelligence agencies should look at how good tools are. Some are very basic.

Monday afternoon, 2:00pm – 3:30pm: Authentication and Crypto

(1) John White, "Securing the Grid, EGEE to EMI and beyond"

abstract: The Grid computing model has been used by various research communities, especially High Energy Physics, to obtain large and reliable computing resources. The Grid uses a trust model that has been developed over many project years. This model strives to ensure that the Grid users and resource owners can be confident in each other's provenance and motives. This trust model is necessary for resource owners to allow the Grid middleware to be installed and run.

The EMI project combines the three most popular Grid middleware in Europe. The EMI Security infrastructure, that encompasses the ARC, gLite and UNICORE models, has been deployed in the first two years of the project. The overall strategy to evolve this infrastructure has been planned and the needed libraries and services developed. During the last year of the EMI project these additions and further strategy refinements will be made.

In order for all EMI services to handle authentication in a common manner a set of specifications for authentication libraries has been made. These libraries are now being produced in C, C++ and Java. A common SAML profile has been specified for common expression of Authorization attributes across the middleware. In order for authorization schemes to be written a common XACML profile has been produced and integrated to EMI some services (e.g. Argus and CREAM). A general request from Grid users for less credential handling and more usage of institutional credentials is leading to the development of a Security Token Service that eases access to Grid services across the middleware stack. A common delegation method is being specified and produced in order to consistently transfer proxies in a secure manner. Security conscious users have the possibility to use

EMI services through pseudo-anonymous identities and encrypted data storage.
These Grid middleware services may also be used to extend the trust model into Cloud-type resources. An example of such a deployment is shown.

Q&A

How is the grid different than the cloud?

In the grid world someone is responsible for what we did in the grid. That's not true for clouds.

Dave and Bart: So the user trusts the site without any evidence?

Yes

The cost of ownership is a question.

Operation cost including electricity is very high. So cloud is affordable at that regard.

What is the optimization of operation in the Grid?

Since computation in physics is huge, Grid is affordable.

Bart - Can you trade off the two (Cloud and Grid) using each other's strengths?
Dave - How much more expensive is it?

About 2.5 to 1 cheaper in the GRID.

Dave: Is it computing or storage in the cloud that's prohibitive?

It's access. If you get charged for every access that gets very expensive.

(2) Shiho Moriai, "Lightweight Cryptography for the Cloud: Exploiting the Power of a Bit Slice Implementation"

abstract: Shiho Moriai gave a talk about lightweight cryptography and its application to Cloud by exploiting the power of a bitslice implementation. This talk is based on the paper by Seiichi Matsuda and Shiho Moriai presented at CHES2012. This paper showed the great potential of lightweight cryptography in fast and timing-attack resistant software implementations in cloud computing by exploiting bitslice implementation. This is demonstrated by bitslice implementations of the PRESENT and Piccolo lightweight block ciphers. In particular, bitsliced PRESENT-80/128 achieves 4.73 cycles/byte and Piccolo-80 achieves 4.57 cycles/byte including data conversion on an Intel Xeon E3-1280 processor (Sandy Bridge microarchitecture). This is about 16% faster than the

previously-known best performance of AES in bitslice implementation. It is also expected that bitslice implementation offers resistance to side channel attacks such as cache timing attacks and cross-VM attacks in a multi-tenant cloud environment. Lightweight cryptography is not limited to constrained devices, and this work opens the way to its application in cloud computing.

Q&A

Bart - Is computation on encrypted data practical using homomorphic operations?

Sometimes.

Dave - Can you take advantage of hardware accelerators?

Yes, you may use GPU of MMX or any other.

Bart - why the focus on light weight encryption at the servers.

A: motivation for light weight is on the collecting sites.

Bart - Why do you bother with hardware acceleration on the server side?

A: We don't. The focus on hardware acceleration is on the client side.

Dave - You assume the secret in the sensor for decrypting key data. Isn't there a problem with key distribution?

Bart - Can you employ data chaining with bit sliced encryption?

Keeping the order is difficult.

Comment: Google is pushing SSL, heavyweight but it can be affordable in terms of network delay.

Comment: With SSL the handshake is a problem.

We will do some prototyping. Just encryption time is not a measure. But networking is a bit slow.

Loren: Google says that the cost of the network is becoming so high that the cost of public key is in the noise.

Q: Can we use this encryption for vm migration?

Yes.

Yoshio - Can this be used for streaming?

Bart - You need retransmission, does that work? Do you need a reliable bit protocol?

A: We use it for less frequent data.

Monday afternoon, 4:00pm – 5:30pm: Protocols

(1) Eisaku Sakane, “A Study of On-line Interactions between Public Key Infrastructure Components”

abstract: Since production-level operation is heavy duties it is significant to make operation of certificate authority (CA) more efficient and decrease the manual work of the CA operators. Among CA operations, we consider certificate renewal in which the distinguished name of certificate remains unchanged but the key-pair is newly generated. We especially present the problems of certificate renewal based on NAREGI-CA implementation; however, it is probable that same problems arise in other implementation. We propose an improvement of the on-line interaction between Public Key Infrastructure components to solve the problem, implement new functions, and discuss our proposed method. The proposed method makes on-line certificate renewal process more secure.

Q&A

Yoshio - What does “cost” mean?

Operation complexity, not money.

Yoshio - You should declare why the revocation is necessary. This must be depending on Naregi.

Sure, this depends on Naregi circumstances.

Linda - Many organizations are doing similar efforts. How about joint activity for that?

Yoshio - Any comparison with other protocol?

Current advantage of Naregi-CA is routine of renewal period judgment.

(2) Akira Otsuka, “e-Voting over Clouds: Unconditional Security against an Dishonest Majority”

abstract: Unconditional security of e-voting protocol is introduced which are not based on standard computational assumptions, but rely on secure distribution of private keys in the set-up phase from trusted initializer. It requires huge amount of storage and communication, but we can enjoy the strong unconditional, or information-theoretic, security. The introduced e-voting protocol has security properties of privacy of ballots and verifiability, where each voter can check that his vote is counted in the final tally.

Q&A

Barton - What is "unconditional Security"?

It is a technical word meaning "no computational assumption".

Barton - Are you saying only one key needs one-year supercomputer power?

Yes.

(Comment by Shiho): The same condition for other Public Key.

Q: What each column means?

Each voter's vote

C: Secret sharing, brief explanation please.

Q: Is this published?

In LNCS 6000

Tuesday, October 16th

Tuesday morning, 9am – 10:30am: Privacy

(1) Gene Tsudik: "Hummingbird: Privacy at the Time of Twitter"

abstract: In the last several years, micro-blogging Online Social Networks (OSNs), such as Twitter, have taken the world by storm, now boasting over 100 million subscribers. As an unparalleled stage for an enormous audience, they offer fast and reliable centralized diffusion of pithy tweets to great multitudesof information-hungry and always-connected followers.

At the same time, this information gathering and dissemination paradigm prompts some important privacy concerns about relationships between tweeters, followers and interests of the latter. In this talk, we assess privacy in today's Twitter-like OSNs and describe an architecture and a trial implementation of a privacy-preserving service called Hummingbird. It is essentially a variant of Twitter that protects tweet contents, hashtags and follower interests from the (potentially) prying eyes of the centralized server. We argue that, although inherently limited by Twitter's mission of scalable information sharing, this degree of privacy is valuable and viable. We demonstrate, via a working prototype, that Hummingbird's additional costs are tolerably low. We also sketch out some enhancements that might offer better privacy in the long term

Q&A

Stefano Zatti - Why is a phantom user (created by the server) different from a legitimate user?

A phantom user can actually extract information from a legitimate user without leaving trace of existence.

Stefano Zatti - Why did you say it does not quite work?

It does actually show to whom the hash has gone (to their pseudonyms) it does not learn who is behind, it cannot be perfect.

Stefano Zatti - Earlier you said it doesn't work, you made us skeptical, and now you have given an argument to be reasonably secure, so what is that doesn't work?

It could be done better. If you want to retain centralized nature, it has to learn that a given hash tag has to go to these 500 users. It will learn the pseudonym behind account. Can achieve kind of privacy only a given hash tag and given message is matched to specific subscriptions. It doesn't learn who tweeted, what's in a message, what's in a hash tag, but a hash tag matches is the limit.

(2) Loren Kohnfelder, "Cloud computing privacy perspectives"

abstract: Describes the Google privacy effort, its guiding principles, and the core methodology we use. In-depth description of the authorization and auditing system in use to Google for fine-grained user data protection will be presented.

Reference to paper:

http://www.cs.berkeley.edu/~dawnsong/papers/2012 Cloud Data Protection for the Masses.pdf

Google Privacy principles:

http://www.google.com/policies/privacy/principles/

Q&A

Bart - Difference between privacy and security?

Yes. Data might not be private, but might want to protect identity usage information.

Bart - Are you confident that all traces are removed when a user says delete account or data including off-line backups?

Yes mostly. Content is deleted or keys destroyed of encrypted data, there may be small traces that data existed, but data itself is gone.

Q. Token lifetime?

It is determined by the owner of the service. It is published internally.

Elisa - Size of log growing quickly, how is this handled?

Long lived batch activities have long token life. Log size is not a problem, plenty of storage.

Q. Why not an option for Google not to keep a persistent copy of encryption key?

Thought about it, but big reason not to is Google can't index content off-line which improves user experience.

Shimojo - How do you protect the results on searches?

Crawled web corpus and index is not encrypted.

Moriai - Is log data encrypted?

It is access-controlled, but not encrypted (will be soon)

Yoshio - How is this architecture implemented?

All home built proprietary. No details, but will send a paper with more information on the implementation.

Linda - How many staffs have access to logs? Do you do vetting on them?

The legal department and a tiny part of security teams that investigate intrusions, whereas the teams have access only to their own product data. Number is very small anyway.

Bart - Are the staff vetted?

No, only some background tests, on criminal records.

Tuesday morning, 11am – 12:30pm: Cloud Security

(1) Atsuhiro Goto, "Inter-Cloud Computing for Secure Social Infrastructure: What's Next?"

abstract: Inter-cloud computing technologies will be one of the "technical and commercial fundamentals" for the secure and reliable cloud computing systems applied to the core of social infrastructures. These technologies will include mechanisms enabling on-demand system reconfiguration and service operation across "autonomous" clouds, operated by different cloud providers, and then providing seamless service deployment for end users even when wide-area disaster occurs. One of the Japanese national R&D projects has been developing the inter-cloud computing technologies for three years and now starts contributing global standardization.

The inter-cloud computing technologies are expected to be openly deployed via standardization as well. Global Inter-Cloud Technology Forum (GICTF) promotes these international standardization

activities through cooperation with global standards bodies.

The next cloud evolution, inter-cloud with Big Data era, is discussed from the view point of "Big Data for Security, and Security for Big Data". Whole architecture should be re-examined for the next cloud evolution.

Q&A

Q - How do students finish a PhD in 3 years?

That's the minimum.

Q - Do they migrate (resources) when a disaster strikes?

Computing resources are reassigned.

Q - Is there a full trust relationship between the clouds?

Yes.

Q - Storage model: Databases are shared among providers. How's that achieved?

The storage model is not yet described. It's under discussion.

Q - Each Cloud provider should provide extra space. Isn't it difficult to achieve?

Remote access will be enhanced.

Q - There's no time to copy big data in the event of a disaster.

I agree. We should have replicas in advance. Otherwise, in case of disaster, we have no time to make a copy.

Q – Is there a priority of applications to be executed in case of disaster?

It should be defined after discussions. In case of disaster, resources are limited. So, you have to prioritize applications.

Q - About sensor networks: Comment on that Google has a fleet of autonomous cars. Attack to unlock cars and turn the engine on could cause problems.

Legacy code can be a major security topic.

Related info can be found in http://www.autosec.org/pubs/cars-oakland2010.pdf

Q - Why does Big Data need re-examination of security?

It is needed because the output is confidential.

David Presotto. Why should we trust cloud computing?
Note taker: Akira Otsuka, Elisa Heymann

(2) David Presotto, “Why should we trust cloud computing?”

abstract: Why should we trust the could? What can cloud providers do to earn that trust? This talks describes a cloud provider, shows where vulnerabilities exists and, as an example of a provider, what Google in particular does/is in the process of doing to protect user data and execution. In particular we address the vulnerabilities caused by internal services sharing the same machine, VMs sharing the same resources, logs saved for debugging/auditing/monitoring, user data stored within the cloud's storage, thousands of developers working on the code and operators/reliability engineers with access to the devices and raw machines. We show that , while we can't give absolute guarantees, we can mitigate any vulnerability to a level as good as or better than services running on the user's own computers.

Q&A

Q - Do you data mine your code changes?

There are tools that fix poor code.

Changes are reviewed by someone else.

Q - Do you have a standardized building process?

Yes. There's a single build. No tools are injected in the process. All the packages to be used should be specified. The build is local.

Q - We've seen vulnerabilities in log files of external software. Any comment?

The inside software is far back. It's hard to attack it. Nodes used for debugging are watched.

The code is C++, and secure programing techniques are used.

Q - Binary forensics to detect what language, etc. Would those tools be useful for Google?

They could be interesting.

Q - How is the user authentication work?

Engineers and insiders are one set of authentication.

Every connection machine to machine is doing with public/shared keys generated for the session.

Communication is through RPC. There's an extra control with a server which says which users are allowed on a machine.

External users come in a cookie based way. The cookie once in the system is changed and that links to Loren's talk.

Q - Tracking transformations?

No. But in the case of multiple users getting the same attachment, only one is stored. For users collaborating, multiple ACL are generated.

Q – Are there logs?

There are logs of the process itself. The logs are for debugging and not related to the user data.

Logs for queries, and get evacuated to a log storage system. It's an infinite large append file. There's a tool which takes the data and aggregates it so engineers could process the data on the fly. Engineers have different levels of permissions.

Queries on user data with different ownership make privilege management difficult.

Tuesday afternoon, 1:30pm – 3:00pm: Vulnerabilities and Intrusions

(1) Akihito Nakamura, “Unified Vulnerability Management SaaS Based on Open Standards”

abstract: In this presentation, we discuss how to automate continuous vulnerability assessment and remediation processes based on the open standards and public contents. Also, the design and implementation of vulnerability management system for multi-platform administrative domains are presented. The system collects asset information in the domain and analyzes the up-to-date impacts of vulnerabilities using public security contents. The assessment result is accurate and prioritized based on the standard testing and scoring scheme, and provides data for making informed decision about cost-effective remediation. In addition, the data model and test procedures are well-abstracted for various kinds of common platforms. These features enable the management process to automate and contribute for decreasing the workload and saving time of system administrators. The server components of the system have REST-style Web services APIs and can be easily deployed on cloud in the SaaS layer. This architecture is flexible enough to inter-operate with multi-platform, external, and even future emerging tools and services with rapid provision and scalability.

Q&A

Q: Isn't there need for automation in each of the scan, analysis, remediation and planning steps?

Q: When you say "vulnerability", do you mean a problem with the code of a program? or what is installed on a particular system?

We never scan source code. We are looking for known vulnerabilities.

Q: How do you measure the effectiveness or success of this system?

That's difficult to answer. It depends on the maturity of the administrators. Skilled administrators may not need such a tool, but beginners may benefit

Q: How do you evaluate this?

We can evaluate the performance of this system, but usability is difficult to evaluate.

(2) Takamichi Asou: “An Analytical Study for Sensor Service Provider to Keep their User’s Security”

abstract: Sensor service attracts huge attention these days. It is easy to image sensor service provider would utilize public cloud as computing resources for sensor data management or secondary use of it in near future. However I should be careful to use sensor data, which is sensitive one, because there would be serious situation for personal information leakage and invasion of privacy if I treat those by uncertain ways. my study is what sensor service provider should care when treat sensor data on cloud. In addition I focus to study what is sensor and sensor data? I consider the difference between the scenarios to use cloud. Consequently, I found that treating sensor data for both primary use and secondary use on cloud, I couldn’t find the proper method to keep it secure so far.

Q&A

Q: When you say "security for sensor data”, what is your main goal?

My main concern is information leakage.

Q: You assume that there is an attacker inside the cloud provider?

Yes.

Loren - If you don't trust the provider, they could delete your data, right?

Yes, but right now, we're only worried about leakage.

Q: Can you explain the 3rd scenario (C) on slide 15?

I'm using the cloud as storage, so I need some sort of encryption in the cloud but I have to decrypt it to the cloud. On the other hand, it should be used for secondary use without decryption.

Q: How would you do statistical operations on encrypted data come from multiple users with different keys?

Moriai: You can use re-encryption schemes via a proxy, though they are expensive

Tuesday afternoon, 3:30pm – 5:00pm: Policies and Experience

(1) Linda Cornwall, "Grid security policies, procedures, and activities: defining responsibilities and making things happen"

abstract: This talk describes the security activities being carried out in EGI. It describes the various security groups, their activities, and their interactions. This includes the EGI Security Policy Group writes policies which define expected behavior of sites. The EGI Software Vulnerability Group issue handling process is described: anyone may report a suspected vulnerability in the EGI infrastructure, the vulnerability is investigated by the Risk Assessment team, if valid a Risk assessment is carried out and the target date for resolution set according to the risk. Monitoring of sites for the presence of Critical and High risk vulnerabilities is carried out, and sites instructed to rectify them. While many of the activities are designed to prevent security incidents, incidents do occur and the incident handling procedure is summarized. Another activity, the EGI Security Threat Risk Assessment which was carried out between February and June 2012 is also briefly described.

Q&A

David - How often do you have security group drill?

Usually once a year.

Atsuhiro - How many people do you have in each security group?

For example, we have 30-50 members in CSIRT. These include security officers in each country. Not all members are full time member.

Barton - How many times does CSIRT actively investigate incidents events? Do you summarize lessons to learn through the investigation? Do you think people in the security group are doing well?

The investigation is done once per month. We discuss lessons to learn from the results and evaluate activities in each site. People in the groups work together well, but some groups have lack of people.

Loren - Have you seen targeting attacks before?

Not sure.

David - Your procedure found an incident in the PC and unplugs the PC. Do you think that it is enough?

We do not want spread out of the incident to the grid.

Loren - How do you mitigate the policy?

We need to discuss this issue involving more security people.

Shiho - How did you derive the impact score and how do you update the score?

We have a set of rules for rating the impact score. The updating period is probably 18 months.

(2) Yoshio Tanaka, “Migrating from Grid to Cloud: GEO Grid’s experience from Security Perspective”

abstract: GEO Grid is aiming at providing a Cyber Infrastructure which provides federated access to Earth observation data for Earth sciences. The security of GEO Grid was designed and implemented using Grid Security Infrastructure in which PKI-based authentication and attribute-based VO-level authorization using VOMS were used. Through the use by end users of GEO Grid, we could prove that GEO Grid's security is secure and stable, however, insights gained through the experiments also gave us negative issues of PKI-based security. For example, GSI is less familiar with Web services, hence interoperation with SNS could not be expected. In order to broaden the userbase, we have decided to change GEO Grid's security from GSI to Cloud-friendly security such as OpenID and OAuth. In this talk, I will give the motivation of our work and the design of the security architecture using OpenID and OAuth.

Q&A

Loren - Is capacity of resources assigned to VO is limited?

It depends on the utilization. Currently, we have no consideration about this issue.

Loren - How much latency do you observe to receive query results in the GeoGrid system (in the use-case example in the presentation slide)?

It varies from few seconds to less than one minute.

Shiho - How long is the lifetime of sensors to collect earthquake info?

Not sure, but die before the end.

Loren - How do you decide fee for looking satellite data? If the data owner is the Japanese

government, should they offer data free?

There are needs for make satellite data free, but currently it is not free. Future policy will be discussed by the Government.

Kento - Are computing resources dedicated to QuiQuake jobs? If no, how do you schedule the QuiQuake jobs?

The computing resources are not dedicated. We have not had a sophisticated job scheduling policy, but we launch VMs which are needed to run jobs through Condor.

Atsuhiro - You should not use OAuth for AuthN, should you?

Right. I am working on this issue and expect to use OpenID Connect rather than OpenID and OAuth.

Wednesday, October 17th

Wednesday morning, 9am – 10:30am: New Environments

(1) Stefano Zatti: “Clouds in Space? The approach of the European Space Agency to reach the clouds”

abstract: The talk summarized the basic needs for ESA to acquire or use a cloud based infrastructure. The approach taken has been to create a working group to develop a model for the needs of all the different parts of the organization (Directorates) and an approach towards satisfying them. teh constraints and the legal requirements were addressed in some detail.
Finally, the project Helix-Nebula where esa is involved together with CERN, EMBL, and european industry, was introduced.

Q&A

Loren - Can you explain 10 times cost reduction?

In order to reduce cost, we should make a single contract with a cloud provider within divisions and should not make many contracts.

David - Press?

NASA : no comments

ESA : admit

Barton - How do you communicate with space? If do so, there should be security.

Because of the subject of this grant, space is source of data.

Barton - In U.S., democratic contract is a problem.

It is important data to be public, quickly and in cheap way.

Loren - Security problems may not be solved.

More new protection will be provided in the future.

David - Related to “Regulatory Framework” slide, how do you use (choose) cloud when all these constraints can’t be fulfilled but some are done?

It depends on cloud provider.

The important thing is not to fulfill the constraints but to list them up.

Aida - What do you mean by “use cloud as a grid”?

ESA uses cloud for processing, not storing data, use grid service as SaaS.

Gene - Isn’t it contrary to grid model?

We consider it as “grid of cloud”.

Tanaka - What protocol do you use for data access? There are OGC protocols and others.

I don’t know.

Tanaka - Do you have any collaboration with other projects? We expect collaboration between AIST GEO Grid and ESA.

Yes, we want.

John - How much data need to be encrypted?

Data encryption is EU policy matter.

(2) Shinji Shimojo, “Can the New Generation Networking Idea and its Testbed Help to Improve Security?”

abstract: In this talk, I talk about the recently proposed ideas about new generation networking or future internet and elaborate how these ideas can help to improves current security issue. The important element of future internet includes, network virtualization, programmability on network, ID/Locator Separation, Contents Oriented Network, In Network Processing, etc. Among these ideas, most important elements are the programmability on the network and network virtualization which are realized through Software Defined Network or SDN. I also talk about JGN-X, which is the

testbed network for trying these new ideas in reality.

Prof. Shimojo's Talk

Q&A

Gene - What does "SDN" stands for?

Software-Defined Network

Gene - Named data network, I'm involved in it. Do you know?

Yes. we may collaborate.

Barton - What's the difference between overlay network and SDN?

SDN can define virtual network in the network layer, but overlay network do in application layer.

Barton has tree-based overlay network technique.

Barton - How do you control security and access control of the network?

Security configuration is distributed in each router.

By using SDN, configuration can be centralized.

You can support both SDN and current routing in the same switch.

Loren - How do you distinguish malicious packets?

Packets are sampled and sent to the DPI.

DPI itself is out of scope.

Goto - Who has initiative in future network, IT industry or network industry? How about equipment?

IT industry. Network industries may be legacy. I don't know about equipments.

David - Google is considering adopting SDN.

Loren - Why do you use tunneling?

There are many reasons why use tunneling, not only security.

David - Google usually use tunneling for security.

There are overheads for tunneling.

Wednesday morning, 11am – 12:30pm: Mobile Security and Malware Analysis

(1) Rika Hayashi, “Android Security Improvement by Visualization of Application Behavior”

abstract: Android applications which work out of users' intention,"suspicious application", are becoming from "annoying" to "threats". However, it is very difficult for users to notice such suspicious applications because users can only verify the permissions required by them before installation. The purpose of this study is to provide users with information to judge whether applications are safe or suspicious even after installing them. Here, we propose the way to show users application behavior with two features. One is to inform users that applications take important actions on the spot, and another is to show users the brief history of important actions taken by applications. We have been developing an evaluation system and show that it is effective to add those features to accomplish the purpose.

Q&A

Q - Do you have any user experience or feedback?

Not yet.

Presotto - I, on the other hand, I would love this. Applications do too many things and it would be great if they could show it.

Q - If the user denies the permission, what do you do to the application?

We stop the application. However what she wants to do is, she wants the program to continue to run, but just that part of the application unable to run or be null.

Kohnfelder - I think that there is the Cynogen (sp?) mod for Android that allows you to pick and choose permissions.

Presotto - Right now, when you install an application, permissions are all or none.

(2) Hiroki Hada, “Using a Database in the Cloud for the Static Analysis of Malware”

abstract: It is crucial to analyze malware behaviors precisely and efficiently to clarify the affected extent in the forensic process when our social infrastructure systems are under targeted-attack.

We propose the new analysis system architecture using a database in the cloud which makes static malware analysis effective. One of the key components in this system is a similarity analysis function which compares execution code of the target malware with already known malware in the

database. We evaluated this component with some malware samples and show this architecture is available.

Q&A

Q - When you say penetration test, do you mean network? systems? applications?

Network layer penetration testing.

C - I wouldn't characterize static analysis as more precise and dynamic as less precise. They simply have different limitations.
Kohnfelder - And if you add the "manual" aspect of static analysis, the "perfect" part is even less likely.

Q - If it's written in a high level language, why are you analyzing assembly code?

Malware only provides binary code.

Q - What are malware examples that you are showing?

Spy8

Q - Are they packed?

Yes, but I unpacked them manually.

Q - How do you plan to unpack the malware as a general approach?

Use other people's work. We don't have such tools now.

Q - Is the similarity calculated by hash functions? Must the instructions be identical?

Yes, they must be identical at the basic block level. We identify identical instruction sequences (green), ones with at least one instruction the same (yellow), and ones completely different.

Goto - The tools first compare the graph structure then compare the contents of basic blocks.

Q: What about polymorphic malware? Does your approach capture differences between different polymorphic versions of the same malware?

Yes, they would be treated as different.

Thursday, October 18th

Thursday morning, 9am – 10:30am: Infrastructures

(1) Kento Aida: “Authentication System for High Performance Computing Infrastructure in Japan”

abstract: This talk presents design of the authentication system for the High Performance Computing Infrastructure (HPCI), which is supported by the Ministry of Education, Culture, Sports, Science and Technology in Japan. The presented authentication system enables single sign-on to computers and shared storages on HPCI by utilizing Grid Security Infrastructure (GSI) and Shibboleth. HPCI started production level operation in September, 2012. This talk also show a short demo of our authentication system currently running on HPCI.

Q&A

Loren - What’s advanced software?

Takizawa-san will explain in more detail, but it needs root privileges, etc.

Bart - Do you charge the money for industry users?

We have several categories

Bart - How much do you charge for K?I think that it is $5000-6000/hour for Jaguar, etc.

I don’t know.

David - Does the portal know the users of distributed resources?

Shibboleth redirect authentication for their home institutions.

Yoshio - How are you authenticated by MyProxy server?

It’s already done for the demo. I skipped it.

David - What does outside of Japan mean?

Users who want to use HPCI need submit proposals.

The representative of the proposal is responsible.

Check ID card.

If you live in Japan for more than 6 months or more, it is ok.

Bart - Are there any matters which are not yet clear?

It is related to both technical and operational issues.

This is the first infrastructure which uses PKI for federation.

Easy way to use PKI is one of the issues.

John - Why don't you use SLCS CA? You have custom software, but are you interested in? It is deployed in LCG.

OK. Thanks.

Akira - Is there any security abuse experience?

No. We only have two weeks. I don't expect that.

Yoshio - Any plan for auditing?

We provide self-check list for self assessment.

Loren - Too strong auditing is not a solution. What's the problem?

It takes human and time resources.

(2) Shinichiro Takizawa, "VM Hosting for High Performance Computing Infrastructure in Japan"

abstract: High Performance Computing Infrastructure (HPCI) is a Japan national grid infrastructure that federates supercomputers and shared storages so that researchers can efficiently use these resources. However supercomputing resources are difficult to be used for some research areas, such as operating system, system software and distributed computing, because of their operation policies and architectures. I propose HPCI Advanced Software Development Environment (HPCI-AE) to support these kinds of researches. HPCI-AE is a virtual machine (VM) management system on a distributed environment and it can run VMs on any sites depending on users' requests within a few minutes. Its authentication is linked with HPCI authentication infrastructure and users can single sign-on to HPCI-AE from any supercomputers in HPCI.

Q&A

Yoshio - Is network is only one problem to use public clouds?

It is the biggest problem.

Yoshio - How do you select VM location?

Invocation location is specified by users.

Yoshio - Any plan for metascheduler, etc?

It may be possible to use such technologies.

Yoshio - Are VM hosting resources homogeneous?

Software stack is the same, but hardware is different.

David - Unique hardware has advantages.
Are you limiting VMs only for traditional CPUs?

Yes

Loren/David - Do you virtualize the network?

Network is bridged.

Kento - Use VPN services.

Hypervisor selects network interface.

We allow users to use specific network segments in the application layer.

Yoshio - How much do you need to modify RENKEI-VPE for adapting OpenNebula 3?

A lot, and it is not easy to move to OpenNebula 3.

Yoshio - Do you have any plan to use this infrastructure for HPC applications?

Yes, but need to take time.

Yoshio - You need to reduce overhead of virtualization, especially for network. AIST is happy to collaborate.

4. Summary

This workshop was an first meeting of groups in a couple of ways: cloud and grid security researchers; academic, industrial, and laboratory researchers; and Japanese, European and U.S researchers. As such, it was an important meeting to familiarize each other with the problems, approaches, and solutions in each domain. And, perhaps more importantly, sharing the open problems in each area. To quote one of the cloud security researchers: “I had heard about grid computing, but this is the first chance I had to understand what it meant and the security problems. It is clear that such a first meeting is only possible at a center like Shonan, where you have the time to have in depth discussions and establish the personal contacts on which to build future collaborations.

These discussions are just a first step. Already teams are planning some join grant applications and visits to each other’s facilities. Many attendees went home with a list of papers that they wanted to read to learn more about the work that was presented.

There was uniform agreement from the attendees that their horizons were broadened and they

benefits from this meeting. And there was the hope that there could be a future meeting to continue the discussions started at this meeting, establishing more collaborations, and adding new attendees to this core group.