Crafting a Balance between Big Data Utility and Protection in the Semantic Data Cloud

Yuh-Jong Hu Kua-Ping Cheng Ya-Ling Huang
{hu, 99753025, 99753026}@cs.nccu.edu.tw

Emerging Network Technology (ENT) Lab.
Department of Computer Science
National Chengchi University, Taipei, Taiwan

June-12-2013

International Conference on
Web Intelligence, Mining, and Semantics (WIMS'13)

Motivations

1. How to effectively collect and analyze complex big data, including structured and unstructured, is hot but the related privacy issue does not arise much attention.
2. Statistical Disclosure Control (SDC) for microdata protection has been well-established so this is a good starting point.
3. How to achieve a balance between big data utility and privacy protection through the combination of SDC and Semantic Web techniques?
4. Solving a complex big data utility and protection problem requires a multi-disciplinary approach, including statistics and computer science.

Motivations

1. How to effectively collect and analyze complex big data, including structured and unstructured, is hot but the related privacy issue does not arise much attention.
2. Statistical Disclosure Control (SDC) for microdata protection has been well-established so this is a good starting point.
3. How to achieve a balance between big data utility and privacy protection through the combination of SDC and Semantic Web techniques?
4. Solving a complex big data utility and protection problem requires a multi-disciplinary approach, including statistics and computer science.

Motivations

1. How to effectively collect and analyze complex big data, including structured and unstructured, is hot but the related privacy issue does not arise much attention.
2. Statistical Disclosure Control (SDC) for microdata protection has been well-established so this is a good starting point.
3. How to achieve a balance between big data utility and privacy protection through the combination of SDC and Semantic Web techniques?
4. Solving a complex big data utility and protection problem requires a multi-disciplinary approach, including statistics and computer science.

Motivations

1. How to effectively collect and analyze complex big data, including structured and unstructured, is hot but the related privacy issue does not arise much attention.
2. Statistical Disclosure Control (SDC) for microdata protection has been well-established so this is a good starting point.
3. How to achieve a balance between big data utility and privacy protection through the combination of SDC and Semantic Web techniques?
4. Solving a complex big data utility and protection problem requires a multi-disciplinary approach, including statistics and computer science.

Research Goals

1. How can we provide semantic metadata markup services for structured data to establish a semantic data cloud?
2. How can we provide data integration and protection services within an outsourcing homogeneous data source for effective microdata analysis without fear of illegal data disclosure?
3. How can we apply data exchange and protection services across outsourcing heterogeneous data sources to have effective microdata sharing and analysis without fear of illegal data leakage?
4. How can we design and implement semantics-enabled policy of SDC for data protection while enforcing data analysis?

Research Goals

1. How can we provide semantic metadata markup services for structured data to establish a semantic data cloud?
2. How can we provide data integration and protection services within an outsourcing homogeneous data source for effective microdata analysis without fear of illegal data disclosure?
3. How can we apply data exchange and protection services across outsourcing heterogeneous data sources to have effective microdata sharing and analysis without fear of illegal data leakage?
4. How can we design and implement semantics-enabled policy of SDC for data protection while enforcing data analysis?

Research Goals

1. How can we provide semantic metadata markup services for structured data to establish a semantic data cloud?
2. How can we provide data integration and protection services within an outsourcing homogeneous data source for effective microdata analysis without fear of illegal data disclosure?
3. How can we apply data exchange and protection services across outsourcing heterogeneous data sources to have effective microdata sharing and analysis without fear of illegal data leakage?
4. How can we design and implement semantics-enabled policy of SDC for data protection while enforcing data analysis?

Research Goals

1. How can we provide semantic metadata markup services for structured data to establish a semantic data cloud?
2. How can we provide data integration and protection services within an outsourcing homogeneous data source for effective microdata analysis without fear of illegal data disclosure?
3. How can we apply data exchange and protection services across outsourcing heterogeneous data sources to have effective microdata sharing and analysis without fear of illegal data leakage?
4. How can we design and implement semantics-enabled policy of SDC for data protection while enforcing data analysis?

Contributions

1. Propose concepts of a semantic big data analysis pipeline to enable automated data analysis, protection, and interpretation services.
2. Semantics-enabled policies, as a combination of ontologies and rules, are represented and enforced for big data in the statistical databases.
3. Provide transparent SDC selection techniques for data users on solving a data analysis and protection of the statistical databases.
4. Preliminary results are discovered on crafting a balance between data utility and protection through enforcing semantics-enabled policies.

Contributions

1. Propose concepts of a semantic big data analysis pipeline to enable automated data analysis, protection, and interpretation services.
2. Semantics-enabled policies, as a combination of ontologies and rules, are represented and enforced for big data in the statistical databases.
3. Provide transparent SDC selection techniques for data users on solving a data analysis and protection of the statistical databases.
4. Preliminary results are discovered on crafting a balance between data utility and protection through enforcing semantics-enabled policies.

Contributions

1. Propose concepts of a semantic big data analysis pipeline to enable automated data analysis, protection, and interpretation services.
2. Semantics-enabled policies, as a combination of ontologies and rules, are represented and enforced for big data in the statistical databases.
3. Provide transparent SDC selection techniques for data users on solving a data analysis and protection of the statistical databases.
4. Preliminary results are discovered on crafting a balance between data utility and protection through enforcing semantics-enabled policies.

Contributions

1. Propose concepts of a semantic big data analysis pipeline to enable automated data analysis, protection, and interpretation services.
2. Semantics-enabled policies, as a combination of ontologies and rules, are represented and enforced for big data in the statistical databases.
3. Provide transparent SDC selection techniques for data users on solving a data analysis and protection of the statistical databases.
4. Preliminary results are discovered on crafting a balance between data utility and protection through enforcing semantics-enabled policies.

Semantics-enabled Policies

1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analysis and protection, and rules are used for enforcing the principles of data analysis and protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation for microdata protection.
   - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through Datalog rules.
   - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and users' usage context.
   - Data Releasing Policy (DRP) describes what are available SDC methods with de-identifiable PII are disclosed for analysis but data privacy is preserved.

Semantics-enabled Policies

1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analysis and protection, and rules are used for enforcing the principles of data analysis and protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation for microdata protection.
   - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through Datalog rules.
   - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and users' usage context.
   - Data Releasing Policy (DRP) describes what are available SDC methods with de-identifiable PII are disclosed for analysis but data privacy is preserved.

Semantics-enabled Policies

1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analysis and protection, and rules are used for enforcing the principles of data analysis and protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation for microdata protection.
   - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through Datalog rules.
   - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and users' usage context.
   - Data Releasing Policy (DRP) describes what are available SDC methods with de-identifiable PII are disclosed for analysis but data privacy is preserved.

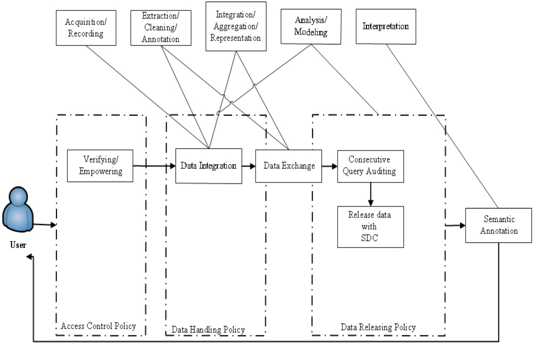
Semantics-enabled Policies

1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analysis and protection, and rules are used for enforcing the principles of data analysis and protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation for microdata protection.
    - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through Datalog rules.
    - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and users' usage context.
    - Data Releasing Policy (DRP) describes what are available SDC methods with de-identifiable PII are disclosed for analysis but data privacy is preserved.

Semantics-enabled Policies

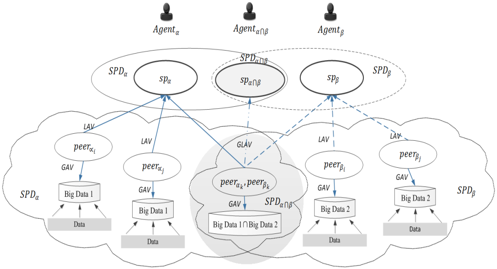
1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analysis and protection, and rules are used for enforcing the principles of data analysis and protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation for microdata protection.
   - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through Datalog rules.
   - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and users' usage context.
   - Data Releasing Policy (DRP) describes what are available SDC methods with de-identifiable PII are disclosed for analysis but data privacy is preserved.

Automated Big Data Analysis Pipeline [32]



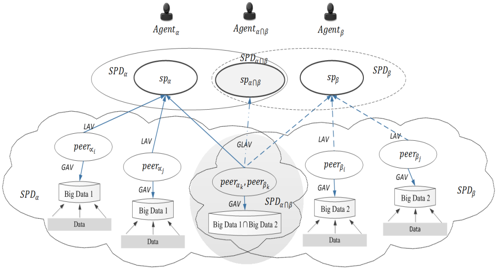
Semantics of Super-Peer Domain (SPD) Cloud

1. Semantics of a super-peer data cloud is described as the policy ontology, including modular concepts of SPD.
2. Semantics-enabled policies perform data integration within an SPD.
3. Semantics-enabled policies are unified to fulfill data exchange across SPDs



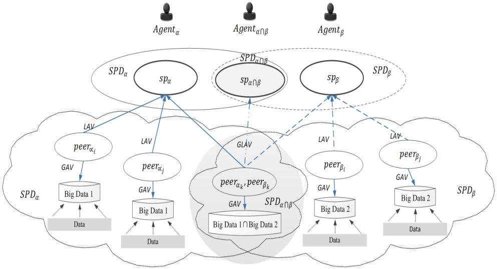
Semantics of Super-Peer Domain (SPD) Cloud

1. Semantics of a super-peer data cloud is described as the policy ontology, including modular concepts of SPD.
2. Semantics-enabled policies perform data integration within an SPD.
3. Semantics-enabled policies are unified to fulfill data exchange across SPDs

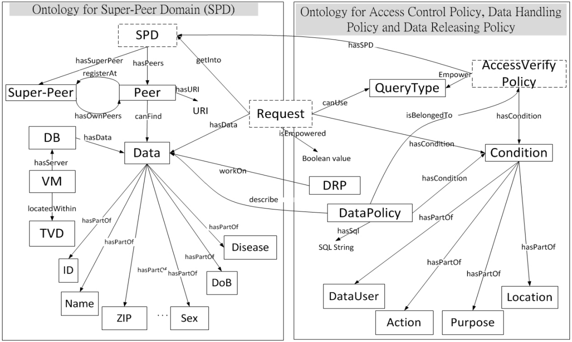


Semantics of Super-Peer Domain (SPD) Cloud

1. Semantics of a super-peer data cloud is described as the policy ontology, including modular concepts of SPD.
2. Semantics-enabled policies perform data integration within an SPD.
3. Semantics-enabled policies are unified to fulfill data exchange across SPDs



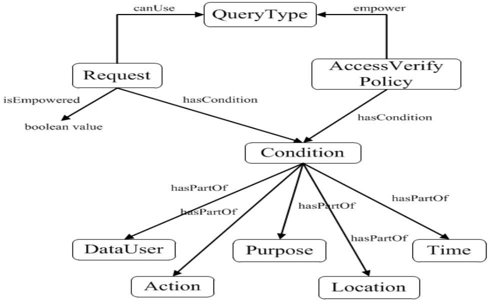
Policy Ontology for Super-Peer Domain Cloud



Ontology for Access Control Policy (ACP)

Definition of ACP Ontology

ACP describes the concept of data usage access control in the super-peer of an SPD.



Rule for Access Control Policy (ACP)

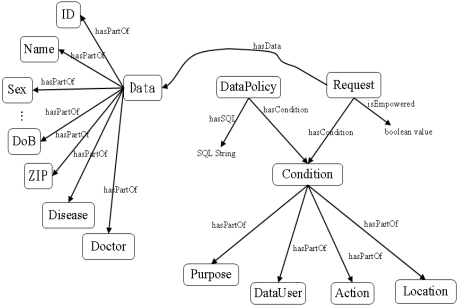
Specification of ACP Rule

$\texttt{Request}(?\texttt{r}) \wedge \texttt{hasCondition}(?\texttt{r}, ?\texttt{c}) \wedge \texttt{Condition}(?\texttt{c})$
$\wedge \texttt{hasCondition}(?\texttt{avp}, ?\texttt{ac}) \wedge \texttt{Condition}(?\texttt{ac})$
$\wedge \texttt{AccessVerifyPolicy}(?\texttt{avp}) \wedge \texttt{sameAs}(?\texttt{ac}, ?\texttt{c})$
$\wedge \texttt{empower}(?\texttt{avp}, ?\texttt{qt}) \wedge \texttt{QueryType}(?\texttt{qt})$
$\longrightarrow \texttt{isEmpowered}(?\texttt{r}, 1) \wedge \texttt{hasQueryType}(?\texttt{r}, ?\texttt{qt}) \longleftarrow (1)$

Ontology for Data Handling Policy (DHP)

Definition of DHP Ontology

DHP describes the concept of semantic metadata markup services and decides which data owners' privacy preferences match which data users' usage context.



Rule for Data Handling Policy (DHP)

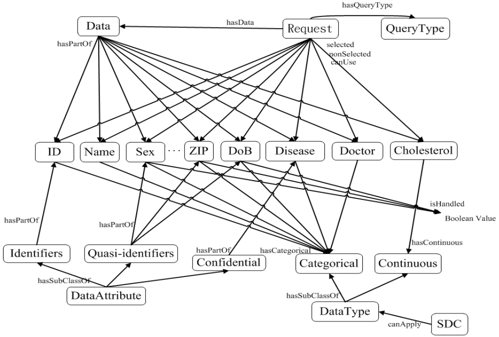
Specification of DHP Rule

$$\texttt{Request(?r)} \wedge \texttt{isEmpowered(?r,1)} \wedge \texttt{hasCondition(?r,?c)}$$
$$\wedge \texttt{Condition(?c)} \wedge \texttt{DataPolicy(?dp)} \wedge \texttt{Condition(?dc)}$$
$$\wedge \texttt{hasCondition(?dp,?dc)} \wedge \texttt{sameAs(?c,?dc)} \wedge \texttt{hasSQL(?dp,?s)}$$
$$\longrightarrow \texttt{sqwrl} : \texttt{select(?s)} \longleftarrow (2)$$

Ontology for Data Releasing Policy (DRP)

Definition of DRP Ontology

DRP describes the concept for which part of PII attributes are allowed to disclose for analysis and still ensures the privacy principles.



Ontology for Data Releasing Policy (DRP)(Conti.)

Definition of DRP Ontology

- hasData.Request(), hasData^{-}.Data().
- hasQueryType.Request(), hasQueryType^{-}.QueryType(PBQs).
- hasPartOf.Data(), hasPartOf^{-}.ID(), hasPartOf^{-}.Name(), ...
- hasPartOf^{-}.ZIP(), hasPartOf^{-}.Cholesterol().
- hasSubClassOf.DataAttribute(),
- hasSubClassOf^{-}.Identifiers(),
- hasSubClassOf^{-}.Quasi − identifiers(),
- hasSubClassOf^{-}.Confidential().
- hasPartOf.Identifiers(), hasPartOf^{-}.ID(id.), ...
- hasPartOf.Confidential(), hasPartOf^{-}.Disease().

Ontology for Data Releasing Policy (DRP)(Conti.)

Definition of DRP Ontology

- hasSubClassOf.DataType(),
- hasSubClassOf^{-}.Categorical(),
- hasSubClassOf^{-}.Continuous().
- hasContinuous.Cholesterol(), hasContinuous^{-}.Continuous().
- hasCategorical.ID(), hasCategorical^{-}.Categorical().
  ...
- hasCategorical.Doctor(), hasCategorical^{-}.Categorical().
- canApply.SDC(generalization), canApply^{-}.Categorical().
  ...
- canApply.SDC(top − coding), canApply^{-}.Continuous().

Rules for Data Handling Policy (DHP)

Specification of DHP Rules

$$
\begin{aligned}
&\texttt{Request}(?\texttt{r}) \wedge \texttt{hasData}(?\texttt{r}, ?\texttt{d}) \wedge \texttt{Data}(?\texttt{d}) \\
&\wedge \texttt{hasPartOf}(?\texttt{d}, ?\texttt{pod}) \wedge \texttt{hasQueryType}(?\texttt{r}, \texttt{PBQ}) \\
&\wedge \texttt{sqwrl} : \texttt{makeSet}(?\texttt{rs}, ?\texttt{pod}) \wedge \texttt{sqwrl} : \texttt{groupBy}(?\texttt{rs}, ?\texttt{r}) \\
&\wedge \texttt{Quasi} - \texttt{identifiers}(?\texttt{qui}) \wedge \texttt{hasPartOf}(?\texttt{qui}, ?\texttt{qpod}) \\
&\wedge \texttt{sqwrl} : \texttt{groupBy}(?\texttt{qs}, ?\texttt{qui}) \wedge \texttt{sqwrl} : \texttt{contains}(?\texttt{rs}, ?\texttt{qs}) \\
&\wedge \texttt{Confidential}(?\texttt{c}) \wedge \texttt{hasPartOf}(?\texttt{c}, ?\texttt{dc}) \\
&\longrightarrow \texttt{sqwrl} : \texttt{selectDistinct}(?\texttt{qui}, ?\texttt{gpod}) \longleftarrow (3)
\end{aligned}
$$

Rules for Data Handling Policy (DHP)(Conti.)

Specification of DHP Rules

$\texttt{Request}(?\texttt{r}) \wedge \texttt{hasData}(?\texttt{r}, ?\texttt{d}) \wedge \texttt{Data}(?\texttt{d})$
$\wedge \texttt{hasPartOf}(?\texttt{d}, ?\texttt{b}) \wedge \texttt{selected}(?\texttt{r}, ?\texttt{b})$
$\wedge \texttt{hasContinuous}(?\texttt{b}, ?\texttt{con}) \wedge \texttt{Continuous}(?\texttt{con})$
$\wedge \texttt{SDC}(?\texttt{sdc}) \wedge \texttt{canApply}(?\texttt{sdc}, ?\texttt{con})$
$\longrightarrow \texttt{sqwrl} : \texttt{select}(?\texttt{b}, ?\texttt{sdc}) \longleftarrow (4)$

Specification of DHP Rules

$\texttt{Request}(?\texttt{r}) \wedge \texttt{hasData}(?\texttt{r}, ?\texttt{d}) \wedge \texttt{Data}(?\texttt{d})$
$\wedge \texttt{hasPartOf}(?\texttt{d}, ?\texttt{b}) \wedge \texttt{selected}(?\texttt{r}, ?\texttt{b})$
$\wedge \texttt{hasCategorical}(?\texttt{b}, ?\texttt{cat}) \wedge \texttt{Categorical}(?\texttt{con})$
$\wedge \texttt{SDC}(?\texttt{sdc}) \wedge \texttt{canApply}(?\texttt{sdc}, ?\texttt{cat})$
$\longrightarrow \texttt{sqwrl} : \texttt{select}(?\texttt{b}, ?\texttt{sdc}) \longleftarrow (5)$

Rules for Data Handling Policy (DHP)(Conti.)

Specification of DHP Rules

$$\begin{aligned}
&\texttt{Request}(?\texttt{r}) \wedge \texttt{hasData}(?\texttt{r}, ?\texttt{d}) \wedge \texttt{Data}(?\texttt{d}) \\
&\wedge \texttt{hasPartOf}(?\texttt{d}, ?\texttt{b}) \wedge \texttt{selected}(?\texttt{r}, ?\texttt{b}) \\
&\wedge \texttt{hasContinuous}(?\texttt{b}, ?\texttt{con}) \wedge \texttt{Continuous}(?\texttt{con}) \\
&\wedge \texttt{SDC}(?\texttt{sdc}) \wedge \texttt{canApply}(?\texttt{sdc}, ?\texttt{con}) \\
&\longrightarrow \texttt{sqwrl} : \texttt{select}(?\texttt{b}, ?\texttt{sdc}) \longleftarrow (4)
\end{aligned}$$

Specification of DHP Rules

$$\begin{aligned}
&\texttt{Request}(?\texttt{r}) \wedge \texttt{hasData}(?\texttt{r}, ?\texttt{d}) \wedge \texttt{Data}(?\texttt{d}) \\
&\wedge \texttt{hasPartOf}(?\texttt{d}, ?\texttt{b}) \wedge \texttt{selected}(?\texttt{r}, ?\texttt{b}) \\
&\wedge \texttt{hasCategorical}(?\texttt{b}, ?\texttt{cat}) \wedge \texttt{Categorical}(?\texttt{con}) \\
&\wedge \texttt{SDC}(?\texttt{sdc}) \wedge \texttt{canApply}(?\texttt{sdc}, ?\texttt{cat}) \\
&\longrightarrow \texttt{sqwrl} : \texttt{select}(?\texttt{b}, ?\texttt{sdc}) \longleftarrow (5)
\end{aligned}$$

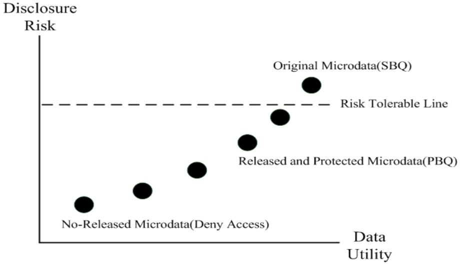
Rules for Data Handling Policy (DHP)(Conti.)

Specification of DHP Rules

$\texttt{Request}(?\texttt{r}) \wedge \texttt{hasData}(?\texttt{r}, ?\texttt{d}) \wedge \texttt{Data}(?\texttt{d})$
$\wedge \texttt{hasPartOf}(?\texttt{d}, ?\texttt{b}) \wedge \texttt{select}(?\texttt{r}, ?\texttt{b}) \wedge \texttt{isHandled}(?\texttt{b}, 1)$
$\wedge \texttt{hasPartOf}(?\texttt{d}, ?\texttt{a}) \wedge \texttt{notSelected}(?\texttt{r}, ?\texttt{a})$
$\longrightarrow \texttt{canUse}(?\texttt{r}, ?\texttt{a}) \wedge \texttt{canUse}(?\texttt{r}, ?\texttt{b}) \longleftarrow (6)$

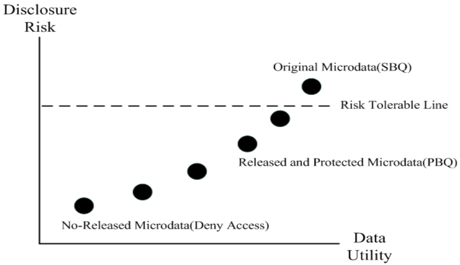
Semantic Data Analysis and Protection

- Improve the situation, where SDC enforcement is obliged to original data providers and a data analytics user lacks the flexibility to choose suitable SDC methods.
- Seek a balance between a data owner's right for privacy protection and a data user's need for data analytics through transparency of SDC methods releasing.
- Semantics-enabled Data Releasing Policy (DRP) calls for which SDC methods and ensures maximum data utility with a tolerable data disclosure risk.



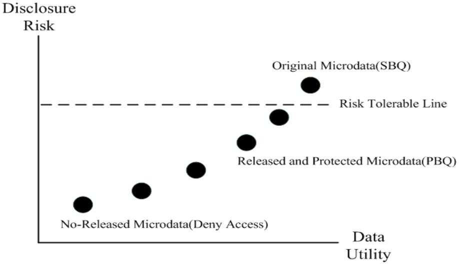
Semantic Data Analysis and Protection

- Improve the situation, where SDC enforcement is obliged to original data providers and a data analytics user lacks the flexibility to choose suitable SDC methods.
- Seek a balance between a data owner's right for privacy protection and a data user's need for data analytics through transparency of SDC methods releasing.
- Semantics-enabled Data Releasing Policy (DRP) calls for which SDC methods and ensures maximum data utility with a tolerable data disclosure risk.

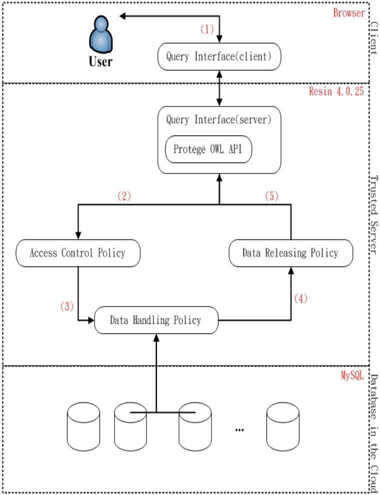


Semantic Data Analysis and Protection

- Improve the situation, where SDC enforcement is obliged to original data providers and a data analytics user lacks the flexibility to choose suitable SDC methods.
- Seek a balance between a data owner's right for privacy protection and a data user's need for data analytics through transparency of SDC methods releasing.
- Semantics-enabled Data Releasing Policy (DRP) calls for which SDC methods and ensures maximum data utility with a tolerable data disclosure risk.



A Three-Tier SDC Prototyping System



Related Work

Major Papers Cited

- Privacy protection for big data:
  [10] [37] [45]
- Statistical Disclosure Control (SDC):
  [1] [12] [16] [29]
- Privacy-aware access control policy:
  [2] [5] [6] [31] [46] [47]

Conclusion and Future Works

Preliminary Results

1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which can ensure the privacy protection principles.
2. Supporting a simple but not yet optimal balance between data utility and protection through policies call for SDC methods.

Future Work

1. Establish a distributed R + Hadoop/MapReduce environment to provide big data deep analysis without violating personal privacy.
2. Design and implement an automated big data analysis pipeline system through Semantic Web Services.
3. The ultimate goal is to craft an optimize balance between data utility and protection in the automated big data analysis life cycle.

Conclusion and Future Works

Preliminary Results

1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which can ensure the privacy protection principles.
2. Supporting a simple but not yet optimal balance between data utility and protection through policies call for SDC methods.

Future Work

1. Establish a distributed R + Hadoop/MapReduce environment to provide big data deep analysis without violating personal privacy.
2. Design and implement an automated big data analysis pipeline system through Semantic Web Services.
3. The ultimate goal is to craft an optimize balance between data utility and protection in the automated big data analysis life cycle.

Conclusion and Future Works

Preliminary Results

1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which can ensure the privacy protection principles.
2. Supporting a simple but not yet optimal balance between data utility and protection through policies call for SDC methods.

Future Work

1. Establish a distributed R + Hadoop/MapReduce environment to provide big data deep analysis without violating personal privacy.
2. Design and implement an automated big data analysis pipeline system through Semantic Web Services.
3. The ultimate goal is to craft an optimize balance between data utility and protection in the automated big data analysis life cycle.

Conclusion and Future Works

Preliminary Results

1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which can ensure the privacy protection principles.
2. Supporting a simple but not yet optimal balance between data utility and protection through policies call for SDC methods.

Future Work

1. Establish a distributed R + Hadoop/MapReduce environment to provide big data deep analysis without violating personal privacy.
2. Design and implement an automated big data analysis pipeline system through Semantic Web Services.
3. The ultimate goal is to craft an optimize balance between data utility and protection in the automated big data analysis life cycle.

Conclusion and Future Works

Preliminary Results

1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which can ensure the privacy protection principles.
2. Supporting a simple but not yet optimal balance between data utility and protection through policies call for SDC methods.

Future Work

1. Establish a distributed R + Hadoop/MapReduce environment to provide big data deep analysis without violating personal privacy.
2. Design and implement an automated big data analysis pipeline system through Semantic Web Services.
3. The ultimate goal is to craft an optimize balance between data utility and protection in the automated big data analysis life cycle.

System Demo and Q&A

(Loading wims13demo.mp4)

R. N. Adam and C. J. Worthmann.
Security-control methods for statistical databases: A comparative study.
ACM Computing Survey, 21(4):515–556, 1989.

A. C. Ardagna et al.
A privacy-aware access control system.
Journal of Computer Security, 16, 2008.

A. P. Bernstein and L. M. Haas.
Information integration in the enterprise.
Comm. of the ACM, 51(8):72–79, July 2008.

M. Bezzi et al.
Modeling and preventing inferences from sensitive value distribution in data release.
Journal of Computer Security, 20:393–436, 2012.

A. P. Bonatti.
Datalog for security, privacy and trust.
In *Datalog 2010*, LNCS 6702, pages 21–36. Springer, 2011.

S. Cabuk et al.
Towards automated security policy enforcement in multi-tenant virtual data centers.
Journal of Computer Security, 18:89–121, 2010.

D. Calvanese et al.
Logical foundations of peer-to-peer data integration.
In *Proc. of the 23rd ACM SIGACT SIGMOD SIGART Sym. on Principles of Database Systems PODS-2004*, pages 241–251, 2004.

D. Calvanese et al.
Data management in peer-to-peer data integration systems.
Global Data Management, pages 177–201, 2006.

D. Calvanese and G. D. Giacomo.
Data integration: A logic-based perspective.
AI Magazine, 26(1):59–70, 2005.

A. Cavoukian and J. Jonas.
Privacy by design in the age of big data, 2012.

S. Ceri et al.
What you always wanted to know about Datalog (and never dared to ask).
IEEE Trans. on knowledge and data engineering, 1(1), 1989.

V. Ciriani et al.
Microdata protection.
In T. Yu and S. Jajodia, editors, *Secure Data Management in Decentralized Systems*, pages 291–321. Springer, 2007.

C. Clifton et al.
Privacy-preserving data integration and sharing.
In *Data Mining and Knowledge Discovery*, pages 19–26. ACM, 2004.

H. L. Cox, F. A. Karr, and K. S. Kinney.
Risk-utility paradigms for statistical disclosure limitation: How to think, but not how to act.
International Statistical Review, 79(2):160–183, 2011.

M. Cox and D. Ellsworth.
Application-controlled demand paging for out-of-core visualization.
In *Proceedings of the 8th Conference on Visualization 97*, pages 235–244, 1997.

J. Domingo-Ferrer et al.
Risk-utility paradigms for statistical disclosure limitation: How to think, but not how to act - discussion: A science of statistical disclosure limitation?
International Statistical Review, 79(2):184–197, 2011.

C. Dwork.
Differential privacy.
In *Proc. of the 33rd International Colloquium on Automata, Languages and Programming (ICALP)*, LNCS 4052, pages 1–12, 2006.

C. Dwork.
A firm foundation for private data analysis.
Communications of the ACM, 54(1):86–95, 2011.

A. Eberhart et al.
Semantic technologies and cloud computing.
In D. Fensel, editor, *Foundations for the Web of Information and Services*, pages 239–251. Springer, 2011.

T. Eiter et al.
Rules and Ontologies for the Semantic Web.
Springer, 2008.

R. Faigin et al.

Data exchange: Semantics and query answering.
Theoretical Computer Science, 336(1):89–124, May 2005.

S. Foresti.
Preserving Privacy in Data Outsourcing.
Springer, 2011.

P. Haase et al.
Semantic technologies for enterprise cloud management.
In *International Semantic Web Conference 2010*, pages 98–113, 2010.

Y. A. Halevy.
Answering queries using views: A survey.
The VLDB Journal, 10(4):270–294, 2001.

Y. J. Hu et al.
Semantic legal policies for data exchange and protection across super-peer domains in the cloud.
Future Internet, 4(4):929–954, 2012.

Y. J. Hu, W. N. Wu, and D. R. Cheng.
Law-aware semantic cloud policies with exceptions for data integration and protection.
In *International Conference on Web Intelligence, Mining and Semantics (WIMS'12)*. ACM Press, June 2012.

Y. J. Hu, W. N. Wu, and J. J. Yang.
Semantics-enabled policies for information sharing and protection in the cloud.
In *Proc. of 3rd Int. Conf. on Social Semantics*, LNCS 6984, Oct. 2011.

Y. J. Hu and J. J. Yang.
A semantic privacy-preserving model for data sharing and integration.
In *International Conference on Web Intelligence, Mining and Semantics (WIMS'11)*. ACM Press, May 2011.

A. Hundepool et al.
Statistical Disclosure Control.
Wiley Series in Survey Methodology, 2012.

A. Inam et al.
A hybrid approach to private record linkage.
In *24th International Conference on Data Engineering (ICDE)*, pages 496–505. IEEE, 2008.

G. Karjoth and M. Schunter.
A privacy policy model for enterprises.
In *15th IEEE Computer Security Foundations Workshop (CSFW)*. IEEE, June 2002.

A. Labrinidis et al.
Challenges and opportunities with big data.
Technical report, Computing Research Consortium (CSR), 2012.

M. Lenzerini.
Data integration: A theoretical perspective.
In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, pages 233–246. ACM, 2002.

J. Madhavan et al.
Web-scale data integration: You can only afford to pay as you go.

In *Proc. of CIDR-07*, 2007.

J. Manyika et al.
Big data the next frontier for innovation, competition, and productivity.
Technical report, McKinsey Global Institute, 2011.

D. Martin et al.
OWL-S: Semantic markup for web service.
Technical report, W3C Member Submission, 2004.

C. A. Mora et al.
Top ten big data security and privacy challenges.
Technical report, Cloud Security Alliance, 2012.

M. Morgenstern.
Security and inference in multilevel database and knowledge-base systems.
In *Proceedings of ACM Special Interest Group on Management of Data*, pages 357–373. ACM, 1987.

A. Nash and A. Deutsch.
Privacy in GLAV information integration.
In *ICDT 2007*, LNCS 4353, pages 89–103. Springer, 2007.

J. M. O'Connor and A. K. Das.
SQWRL: a query language for OWL.
In *OWLED*, volume 529. CEUR, 2009.

R. Popp and J. Poindexter.
Countering terrorism through information and privacy protection technologies.

IEEE Security & Privacy, 4(6):24–33, 2006.

K. Schwab et al.
Personal data: The emergence of a new asset class.
Technical report, World Economic Forum, 2011.

F. J. Sequeda et al.
Survey of directly mapping SQL databases to the semantic web.
The Knowledge Engineering Review, 26(04):445–486, 2011.

L. Sweeney.
K-annonumity: a model for protecting privacy.
International Journal of Uncertainty, Fuzziness and Knowedge Based Systems, 10(5):557–570, 2002.

O. Tene and J. Polonetsky.
Privacy in the age of big data: A time for big decisions.
64 Stanford Law Review Online 63, 2012.

S. D. C. d. Vimercati et al.
Access control policies and languages in open environments.
In T. Yu and S. Jajodia, editors, *Secure Data Management in Decentralized Systems*, pages 21–58. Springer, 2007.

J. D. Weitzner et al.
Creating a policy-aware web: Discretionary, rule-based access for the world wide web.
In E. Ferrari and B. Thuraisingham, editors, *Web and Information Security*, pages 1–31. IGI, 2006.