IT@Intel Executive Summary
Intel IT
Data Anonymization
Cloud Computing
June 2012

Enhancing Cloud Security Using Data Anonymization

Data Anonymization

- Can enhance security of data stored in public clouds
- Still allows for useful analytics and reporting
- Warrants further research, including developing further use cases, educating potential enterprise cloud users about the potential benefits and pitfalls of data anonymization, and documenting existing open source data anonymization applications,

Intel IT is exploring data anonymization—the process of obscuring published data to prevent the identification of key information—in support of our vision of a hybrid cloud computing model and our need to protect the privacy of our employees and customers. We believe data anonymization is a viable technique for enhancing the security of cloud computing.

Achieving Privacy Using Data Anonymization

Although we realize that a 100-percent secure cloud infrastructure is impossible. We are exploring the possibility of anonymizing data to augment our cloud security infrastructure. Data anonymization makes data worthless to others, while still allowing Intel IT to process it in a useful way.

Several formal models of security can help improve data anonymization, including k-anonymity and l-diversity.

- k-anonymity attempts to make each record indistinguishable from a defined number (k) of other records. For example, consider a data set that contains two attributes: gender and birthday. The data set is k-anonymized if, for any record, k-1 other records have the same gender and birthday. In general, the higher the value of k, the more privacy is achieved.
- l-diversity improves anonymization beyond what k-anonymity provides. The difference between the two is that while k-anonymity requires each combination of quasi-identifiers to have k entries, l-diversity requires that there are l different sensitive values for each combination of quasi-identifiers.

Other data anonymization techniques include adding fictitious records to the data, hashing, truncation, permutation, and value shifting, just to name a few.

Proof of Concept – Anonymizing Event Log Data

We conducted a proof of concept (PoC), in which we used data anonymization to protect event logging data stored by a public cloud-based SaaS log management supplier. The PoC

was successful in demonstrating that data anonymization can work and that obscured data is still useful for analysis. We were able to perform both performance analysis and security analysis on the anonymized data.

- During the security analysis testing, we didn't detect any active probing of the monitoring VMs. However, we searched older logs and found that there had been probes on the web server. This confirmed our theory that the approach we took in looking for security business intelligence events could detect real events.
- Although the SaaS log management supplier we used during the PoC didn't support number-crunching analytics, such as calculating averages, we were able to pinpoint other performance issues. For example, we discovered that one web site performed two redirects before the user accessed the actual content, thereby increasing the access time.

Although more research is necessary before it is ready for production use, data anonymization can ease some security concerns, allowing for simpler demilitarized zone and security provisioning and enabling more secure cloud computing. We plan to explore data anonymization further, including conducting a more extensive PoC, developing further use cases for data anonymization, educating potential enterprise cloud users about the potential benefits and pitfalls of data anonymization, and documenting existing open source data anonymization applications.

You can find a full discussion of our work with data anonymization and the PoC at "Enhancing Cloud Security Using Data Anonymization."

For more information on Intel IT best practices, visit www.intel.com/it.

This paper is for informational purposes only. THIS DOCUMENT IS PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, SPECIFICATION OR SAMPLE. Intel disclaims all liability, including liability for infringement of any proprietary rights, relating to use of information in this specification. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted herein.

Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and other countries.

* Other names and brands may be claimed as the property of others.
Copyright © 2012 Intel Corporation.
All rights reserved. Printed in USA Please Recycle