Security issues in nosql databases

Security Issues in NoSQL Databases Lior Okman Deutsche Telekom Laboratories at Ben-Gurion University, Beer-Sheva, Israel Nurit Gal-Oz, Yaron Gonen, Ehud Gudes Deutsche Telekom Laboratories at Ben-Gurion University, and Dept of ComputerScience, Ben-Gurion University, Beer-Sheva, Israel Jenny Abramov Deutsche Telekom Laboratories at Ben-Gurion University and Dept of Information Systems Eng. Ben-Gurion University, Beer-Sheva, Israel Abstract—The recent advance in cloud computing and distributed web applications has created the need to store large amount of data in distributed databases that provide high availability and scalability.

In recent years, a growing number of companies have adopted various types of non-relational databases, commonly referred to as NoSQL databases, and as the applications they serve emerge, they gain extensive market interest. These new database systems are not relational by definition and therefore they do not support full SQL functionality. Moreover, as opposed to relational databases they trade consistency and security for performance and scalability. As increasingly sensitive data is being stored in NoSQL databases, security issues become growing concerns. This paper reviews two of the most popular NoSQL databases Cassandra and MongoDB) and outlines their main security features and problems. Index Terms—NoSQL; Security; Cassandra; MongoDB; I. INTRODUCTION The recent advance in cloud computing and distributed web applications has created the need to store large amount of data in distributed databases that provide high availability and scalability. In recent years, a growing number of companies have adopted various types of non-relational databases, commonly referred to as NoSQL databases and as the applications they serve emerge, they gained

extensive market interest. Different NoSQL databases take different approaches.

Their primary advantage is that, unlike relational databases, they handle unstructured data such as documents, e-mail, multimedia andsocial mediaefficiently. The common features of NoSQL databases can be summarized as: high scalability and reliability, very simple data model, very simple (primitive) query language, lack of mechanism for handling and managing data consistency and integrity constraints maintenance(e. g. , foreign keys), and almost no support for security at the database level. The CAP theorem introduced by Eric Brewer [1], refers to the three properties of shared-data systems namely data onsistency, system availability and tolerance to network partitions. The theorem [2] states that only two of these three properties can be simultaneously provided by the system. Traditional DBMS designers have prioritized the consistency and availability properties. The rise of large web applications and distributed data systems, makes the partition-tolerance property inevitable, thus imposing compromise on either consistency or availability. The main promoters of NOSQL databases are Web 2. 0 companies with huge, growing data and infrastructure needs such as Amazon andGoogle. The Dynamotechnologydeveloped t Amazon [3] and the Bigtable distributed storage system developed at Google [4], have inspired many of today’s NoSQL applications. In this paper we analyze the security problems of two of the most popular NoSQL databases, namely: Cassandra and MongoDB. Cassandra [5] is a distributed storage system for managing very large amounts of structured data spread out across many commodity servers,

while providing highly available service with no single point offailure. Cassandra aims to run on top of an infrastructure of hundreds of nodes. At this scale, components fail often and Cassandra is designed to survive these failures.

While in many ways Cassandra resembles a database and shares many design and implementation strategies therewith, Cassandra does not support a full relational data model; instead, it provides clients with a simple data model that supports dynamic control over data layout and format. Cassandra was designed to support the Inbox search feature of Facebook [6]. As such it can support over 100 million users which use the system continuously. MongoDB [7] is a document database developed by 10gen. It manages collections of JSON-like documents. Many applications can thus model data in a more natural way, as data can e nested in complex hierarchies and still be query-able and indexable. Documents are stored in collections, and collections are in turn stored in a database. A collection is similar to a table in relational DBMS, but a collection lacks any schema. MongoDB also provides high availability and scalability by using Shardings and Replica sets (see below). The increasing popularity of NoSQL databases such as Cassandra and MongoDB and the large amounts of userrelated sensitive information stored in these databases raise the concern for the confidentiality and privacy of the data and the security provided by these systems.

In this paper we review the main security features and problems of these two database systems. We start with a brief overview of Cassandra and MongoDB functionality in section II. We then discuss security features of

Cassandra and MongoDB in sections III and IV respectively. We conclude in section V. Since much of the discussion is based on open-source Internet documents, it naturally reflects the situation at the time this paper is written 2011 International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11 978-0-7695-4600-1/11 $26. 00 © 2011 IEEE DOI 10. 1109/TrustCom. 2011. 70 541