

## Security Challenges for Rich-Media Educational Environments

Robert Grimm

*New York University*

rgrimm@cs.nyu.edu

Medicine is undergoing a major and growing chasm between scientific knowledge and medical practice. On one side, rapid advances in molecular biology are reshaping medical science. On the other side, managed care has resulted in drastically reduced lengths-of-stay in hospitals and a general compartmentalization of medical practice. As a result, it is becoming increasingly difficult to train physicians that can provide state-of-the-art medical care, as medical practitioners cannot keep up with the rapidly changing basic sciences, do not have enough context to make appropriate diagnoses, and may rely on out-dated procedures or drug regimens.

The premise of the Infrastructure for Rich-Media Educational Environments (IRMEE) project at New York University is that a sustainable solution requires the integration of medical knowledge across specializations, between theory and practice, and across geographical boundaries and time. The chosen approach is to create a web-based rich-media environment that (1) provides ubiquitous and lifelong access to educational and scientific materials, (2) structures educational content along narrative lines to re-establish missing context, and (3) fosters a community of students and practitioners not bound by geography. Experiences at NYU's medical school with a set of prototypes support the general approach, demonstrating that rich-media educational environments have advantages over textbooks, educational videos, and lectures alike.

Unfortunately, the straight-forward multi-tier web architecture used for these prototypes has serious scalability constraints and does not provide an adequate basis for realizing the larger vision of IRMEE. To overcome this major deficiency, we are building a more scalable content delivery infrastructure. The goal is to combine the usability of familiar web content management systems with the scalability of peer-to-peer content distribution networks (CDNs) built on distributed hash tables. We aim to achieve this goal by allowing for the execution of application-specific services, which are expressed through scripts, within the content distribution network instead of only on the server. Our architecture leaves both clients and servers unchanged, thus letting us track any advances in web functionality. Furthermore, its scripting-based programming model is already familiar to web developers, thus significantly reducing the barrier to entry in developing applications.

While we believe that a scripting-enhanced CDN provides an appropriate solution for scaling IRMEE, our architecture also raises two important security challenges. First, since the CDN is implemented as a peer-to-peer system, content integrity becomes an important issue. Without additional safeguards, CDN nodes can modify or replace content with their own, arbitrary versions.

Some replaced content may be obvious—consider spam-like advertisements—but other content may be considerably less obvious and consequently more dangerous—consider falsified medical research reports. To make matters worse, established solutions for ensuring content integrity, such as cryptographic hashes, are ineffective in our architecture, as scripting-enabled CDN nodes, by definition, may modify or even create content.

Second, some content, such as students' contributions to discussion groups, may refer to actual patients' case histories. As medical data must be kept private, access to user-generated content should be restricted to authorized users. However, as the peer-to-peer CDN is generally untrusted, authorization can only be performed by the original servers. The simplest solution is to partition all content into two categories, public and private. Public content will be accessible through the CDN, while private content can only be accessed directly through the corresponding server, which is protected through SSL and proper authentication. However, this solution also has the disadvantage of eschewing any scalability advantages of the CDN for private content. Overall, from a security perspective, the issue is to provide strong security guarantees for a relatively untrusted network, while also remaining compatible with the existing web-based infrastructure.