Advances in networking have connected mission-critical applications such as N-tier enterprise systems for e-commerce with open information sources around the world. Problems such as spam in various communications media (e.g., email, web, instant messaging, social networks, voice over IP) and automated click fraud in e-commerce have become economically sufficiently important to make them self-sustaining and growing. We call them Quality of Information (QoI) attacks. QoI problems are growing threats for new applications such as cyber physical systems (CPS) such as environmental monitoring, critical physical infrastructure protection, multi-modal integrated transportation, and supervisory control and data acquisition (SCADA) applications, which would cause serious damages if suffering from QoI attacks. QoI problems are inherent to open communications media such as Internet since new connections are accepted by default. Further, QoI attacks do not require any system or network malfunction (e.g., buffer overflow) to succeed, since they feed polluted or deceptive information to victim applications through legitimate APIs or user interfaces.

QoI attacks are already a significant practical problem. MAAWG (Message Anti-Abuse Working Group) reports show about 80% of billions of email messages flowing through the backbone are spam. Similarly, web spam pages have grown to a significant percentage of total web pages (between 13.8% and 22% of all Web pages). Financial damages caused by phishing (an online form of pretexting to gain sensitive information, usually through spam emails) were estimated at $3.2B in 2007. Given this kind of financial incentive, the underground economy (i.e., organized crime) is expected to continue to finance and expand these attacks. We call QoI attacks Denial of Information (DoI) attacks, which are the information analogs of Denial of Service (DoS) attacks. Given the prevalence of distributed DoS attacks even today, it is clear the DoI problem will represent a long term challenge.

Although significant progress has been made in the combat against spam (particularly in the email and web media), the problem has persisted due to the increasing sophistication of spam content and tools. One of the fundamental challenges in defending against QoI attacks is an “arms race” between attack and defense, called adversarial classification or adversarial learning. In adversarial classification, adaptive defense techniques learn about new attacks through statistical or machine learning techniques. As defense tools are developed and deployed to identify QoI attack content and protocol traces, attackers study the defenses and attempt to bypass them in the next generation attacks. Generally, attackers may feed targeted information to mislead the learning algorithms. As a concrete example, if defensive tools rely on identification markers that are easily changed or disguised, then an attacker may get around defenses with relative ease. Consequently, QoI attacks such as email and web spam are difficult to extinguish.

We have been collecting large spam data sets from real world applications, including email spam, web spam, social network spam, phishing messages, and spam in other communications media. Using these data sets, we have conducted evolutionary studies that identify spam construction techniques and markers that change over time. These studies show that the “low-entropy” spam
markers (the ones that distinguish the spam messages clearly from the legitimate messages with minimal false positives and false negatives) are typically technical flaws (unintentional bugs) in the spam generation programs (e.g., a header that contains a very specific network protocol identifier automatically inserted by the compiler that created the binary of their spam generation tools). These flaws are relatively easy to fix (e.g., by recompiling their tools and turning off the generation of the header). Consequently, rule-based spam filters tend to have a very large number of rules that only apply to older spam messages. There are many other spam defense methods (e.g., white lists and black lists, source identification protocols, and others) that have been proposed. The prevalence of spam in every open communications media shows the insufficiency of these proposed methods.

While we want to point out the importance of DoI attacks and the need for more research on DoI defense, we also want to mention some positive contributions made by DoI techniques. First, encryption in general and steganography in particular are early examples of DoI technique used to protect information. Second, many privacy protection mechanisms (e.g., k-anonymity in mobile computing) can be considered examples of DoI techniques.

In summary, issues of quality of information (QoI) and denial of information (DoI) have the potential to affect data and applications security in significant ways. They are also inherent problems in open communications environments such as the Internet. We have outlined significant threats in many communications media and mentioned some positive uses of DoI techniques such as encryption and privacy. More research on QoI and DoI issues will be needed for the next generation of secure large scale distributed applications.