Abstract:
Designing a system to provide unidirectional security to current IM systems without the added complication of encryption utilizing Confusion.

IM Systems:
Usual server-client architecture (Jabber protocol is used for this project) looks like the following:

In the case of Jabber all clients and servers are XMPP clients and XMPP servers, which stands for Extensible Messaging and Presence Protocol (the formal name for the Jabber Protocol).

The Jabber protocol relays the messages using XML with special tags for different operations. The messages themselves are sent between the <body> and </body> tags.

Legacy System:
• Offers little in terms of IM (Gaim is used for this project) security options.
• Some systems offer encrypted traffic but they require a bilateral cooperation.
• Additional external security options are difficult to install and use.
• Some current encryption methods have proven to be vulnerable to external attacks.

Confusion:
• Project by Dr. Blaze’s group in the Distributed Systems Lab.
• Allows for unilateral implementation.
• Utilize noise packets to confuse the end receiver.
• The noise packets will be dropped by the network before it gets to the end user because their TTL (time to live) is not high enough. Therefore, would not confuse the end user (in this project’s case the end user can also be the server).

Implementation:
• Two separate programs are written: one to replace the Unix network library and the second to capture the packets via divert socket.
• Libcvore.so is the library that is designed to be preloaded and replace the Unix network library.
• Confuser is the program that captures the packets from the divert socket and send out the noise packets after breaking the packets down.
• Noise packets, each containing 1 word, are sent for every word of the actual text message.

Semantic Noise Generation:
• Messages are English text, thus, the confuse packets utilize the natural structure of the English language to generate noise that is difficult to distinguish via autonomous methods. Thus, the noise must follow the language structure.
• Dadadodo technology created by Jamie Zawinski is used to generate the noise.
• Dadadodo constructs a Markov model based a pre-existing English text.

Security:
• The noise words allow for different combination of possible sentences to be formed
• The maximum number of valid combination of text is $\binom{m+1}{2}$
• However some combination might be able to be eliminated by autonomous methods.
• Lower bound of the combination of noise is at least $C+1$

Conclusion:
While there is much additional research that needs be done to complete the security analysis on the confusion implementation over IM, the current implementation achieved in offering unidirectional security over legacy IM systems. The similar implementation can also be used for email and other forms of network traffic that transport English text.

Cost:
Consider a Jabber IM message to be size $m + x + n$
• $m$ is the size of the actual text message.
• $x$ is the xml overhead of the jabber protocol.
• $n$ is the network overhead (basically everything other than data the packet carries).
• Let $C$ be the number of noise packets generated per word.
• Let $T$ be the time it takes to send one packet.

Number of total packets send from 1 Jabber IM message is
$$4 + \frac{m}{6} + C \times \frac{m}{6}$$

Which implies latency of
$$T \times \left(4 + \frac{m}{6} + C \times \frac{m}{6} + x + \frac{m}{20} + (C+1)\right)$$

And byte cost (number of bytes sent with noise over the number of bytes sent in original message) is
$$\frac{n \times \left(4 + \frac{m}{6} + C \times \frac{m}{6} + x + \frac{m}{20} + (C+1)\right)}{n + x + m}$$

Conclusion:
While there is much additional research that needs be done to complete the security analysis on the confusion implementation over IM, the current implementation achieved in offering unidirectional security over legacy IM systems. The similar implementation can also be used for email and other forms of network traffic that transport English text.