Cyber-attacks are becoming more intense and more sophisticated every day, threatening the safety of company systems. On the other hand, discussion of new services with strict personal identification confirmation has started, on the assumption that the private sector will be able to use the Social Security and Tax Number System. There is a demand for reliable personal authentication technology that will also protect personal information. Fujitsu Laboratories Ltd. is developing the latest technologies to resolve these challenges. This report provides an introduction on technology using artificial intelligence (AI) to identify sophisticated cyber-attacks more efficiently and technology for safe encryption of biometric information.

Application of AI Technologies to Detecting Cyber-Attacks

Background

As cyber-attacks increase in intensity, networks from the public and private sectors are being attacked in various ways. In recent years, in particular, concealed amid the large volume of known attack methods such as vulnerability scans and Denial of Service (DoS) attacks, sophisticated attacks such as advanced persistent threats (APTs) have also been carried out. Monitoring and analysis of logs output from network equipment and so forth is considered to be an effective method for detecting these sophisticated attacks. However, as sophisticated attacks occur only rarely, it is extremely difficult to detect an attack by manual analysis of the enormous logs.

Fujitsu Laboratories has developed security log analysis technologies in which AI technologies are used to visualize threats that are difficult to find manually. These technologies enable log analysts to efficiently identify the sophisticated attacks hidden among the large volume of known attacks.

Newly Developed Technologies

To enable identification of sophisticated attacks, Fujitsu Laboratories developed two technologies: the Outlier Structure Learning Technology and Relational Visualization Analysis Technology.

The Outlier Structure Learning Technology analyzes large-scale logs to extract an outlier structure—a small cluster of data with rare values. As sophisticated attacks occur with such low frequency, there is a high likelihood that they will be included in this outlier structure. The Relational Visualization Analysis Technology is used to create a visualization of data clusters having outlier structures. Visualized outlier clusters have characteristics that appear when shown in graphic form depending on the type of attack. These characteristics can be used to categorize data clusters with outlier structures by attack type.

Log analysts can analyze the visualized results and efficiently extract the sophisticated attacks from among the data clusters.

Security Log Analysis Technology Employing AI: Processing Flow

1. Outlier Structure Learning Technology

Outlier Structure Learning Technology focuses on the frequency of data characteristics that appear in logs and extracts small data clusters that have rare characteristics. Conventional technology performed categorization and extraction of the log focusing only on the characteristics that appeared frequently within the log. Consequently, even if attacks occurring in large volumes could be detected, it was difficult to extract small-scale, irregular attacks concealed among them.

Outlier Structure Learning Technology focuses on characteristics that rarely appear in the log, separates data clusters that share characteristics that appear infrequently and repeatedly integrates the multiple data clusters of infrequent characteristics that have been segmented. This enables the extraction of “outlier structures,” which are small-scale data clusters that share rare characteristics within a log.
2. Relational Visualization Analysis Technology

Next, clusters with outlier structures are plotted on a graph to visualize relationality between the data.

Data clusters with outlier structures also contain many known attacks that occur infrequently. Visualizing those attacks revealed similarities in their graphic forms. Focusing on this characteristic enables log analysts to extract the rare graphic form shapes. Data clusters with rare graphic shapes correspond to attacks with characteristics that differ from others, making it easier to identify sophisticated attacks.

Relational Visualization Analysis Technology compares the results of visualization and extracts data clusters with rare graphic forms. This enables the narrowing down of data clusters with a high likelihood of being a sophisticated attack, enabling log analysts to efficiently identify such attacks.

Verification in an Actual Environment Log

Verification was carried out on the developed technology by applying it to a log obtained from an actual environment. Extracting data clusters with outlier structures and comparing the visualization results enabled the extraction of two-to-three data clusters with rare graphic forms. A detailed analysis of these clusters showed they contained a sophisticated attack. Previously, the laboratory had taken about three months to identify this attack. The newly developed technology identified the attack in about one day. This verified that the developed technology enables efficient identification of sophisticated attacks from logs.

Initiatives Going Forward

Currently, this technology is undergoing trial operations under the supervision of Fujitsu Cloud Services. Looking ahead, after refining the analysis accuracy, the technology will contribute to the safe operation and management of Fujitsu Cloud Services.
Newly Developed Technology

We have developed two technologies for application in biocode technology making use of palm veins: technology to protect biometric data using random numbers, and technology to decrypt confidential data using error-correcting codes.

1. Technology to Protect Biometric Data Using Random Numbers

In encryption, biocode converted to a random number is added to confidential data to generate encrypted data, which is registered in the server.

A decryption code is used as the key when decrypting encrypted data. For decryption, the decryption code is converted into secure data by the device and sent from the device to the server. The decryption code is generated by first converting the biocode using a random number. As the system can randomly select different random numbers for encryption and decryption, a different, secure decryption code can be generated every time.

2. Technology to Protect Biometric Data Using Error-Correcting Codes

Variations in one’s motion or position when inputting biometric data, and the use of a different random number each time can generate slight discrepancies. The discrepancies can be absorbed because they are converted using an error-correcting code. Error-correcting code is widely used as a supplementary technology for data loss generated during transmission.

The decryption process using error-correcting code involves two steps. In step 1, the discrepancies caused when calculating the decryption biocode for the encrypted biocode are corrected. In step 2, the discrepancies caused by calculating the random number used for decrypting the random number added during encryption are similarly corrected, enabling recovery of the confidential data.

In this way, as the biometric data input for encryption and decryption are sufficiently similar, so long as they are both from the same person, the confidential data can be retrieved from the encrypted data using error-correcting technology.

Effect

With the ability to use biometric data for encryption and decryption, the cryptographic key management that had been needed for existing encryption technologies becomes unnecessary. Consequently, as servers holding encrypted data do not have the encryption data at the same time, they can be operated more securely. Moreover, as biometric data used during encryption and decryption is converted using random numbers, the unconverted biometric data will not flow into the network. This means that the use of encryption technology using biometrics can now be expanded to cloud services.

Future Plans

With this technology, we aim to continue improving the speed of decryption processing and expanding the types of data that can be encrypted, while also examining the technology’s applicability to a number of potential use cases, such as the Social Security and Tax Number System. We will also examine ways to develop biocode utilizing the feature of the person encrypting being the only person capable of decrypting, and work to expand the types of applicable biometrics, such as fingerprints.