An R&D Strategy for DFFT

Free flow of data across nations around the world will surely bring about huge benefits to humankind. But there are concerns that it can infringe on an individual's human rights, take economic advantages away from those having more data, or can endanger national or regional security. To challenge these issues, Data Free Flow with Trust (DFFT) was proposed by the prime minister of Japan at the G20 summit meeting in 2019. It emphasizes the concept of trust in data communication. We first consider the difference between trust and security. Then we make a scientific model of DFFT where trust is defined in the context of data flow. Using this model, we list the set of core technological problems for realizing DFFT and introduce the R&D strategy of Hitachi in approaching them.

Haruo Takeda
Akira Ishikawa
Tadashi Kaji
Kenta Takahashi
Toshiaki Suzuki
Tatsuya Teshima
Kyoko Yamamoto
Hiromitsu Kato
Seishi Hanaoka
Tatsuhiko Kagehiro
Shinji Nishimura
Norihiro Suzuki

1. Introduction

Free flow of data across nations around the world will surely bring about huge benefits to humankind. Solving viral infection problems is one current conspicuous example where this is a requirement. But there are concerns that data free flow can infringe on an individual's human rights, take economic advantages away from those having more data, or can endanger national or regional security. To solve these issues, social sciences are taking up the challenge of making international rules, while technological sciences are accelerating development of innovative technologies.

Data Free Flow with Trust (DFFT) was proposed by the prime minister of Japan at the G20 summit meeting held in Japan in 2019. In May 2020, the World Economic Forum (WEF) responded by publishing a white paper entitled, "Data Free Flow with Trust (DFFT): Paths towards Free and Trusted Data Flows"[1]. Mr. Hiroaki Nakanishi, Executive Chairman and Executive Officer, Hitachi, Ltd., contributed to it as a steering committee member. Dr. Akira Ishikawa, a co-author of this article, sat on its working committee. The WEF's Centre for the Fourth Industrial Revolution Japan (C4IRJ) will further follow this up by publishing another white paper. It is on the subject of data governance and due to be published in March 2021[2]. Hitachi is serving as a co-author together with the Ministry of Economy, Trade and Industry, Japan. The main contributors from Hitachi are Dr. Tadashi Kaji and Dr. Hiromitsu Kato. Both are co-authors of this article. The Japanese government, meanwhile, has followed up the prime minister's proposal by establishing the Trusted Web Council in the Cabinet Secretariat in October 2020. It will publish a white paper on Trusted Web in March 2021[3]. Dr. Haruo Takeda, the main author of this article, is serving as a member of the Council.

This article introduces the R&D strategy of Hitachi to contribute to realizing DFFT. In the next chapter, we differentiate trust from the term security, which has been widely used in data communication. Then DFFT modeling is tried in the following chapter to involve the engineering science community in this discussion. Using this model, trust is defined in the context of data flow. In the final chapter, we identify the core technological problems for implementing DFFT and introduce the R&D strategy of Hitachi in approaching them.

Figure 1 — Security and Trust

1. Security $\overset{\text{implies}}{\rightarrow}$ trust, Trust $\overset{\text{does not imply}}{\not\rightarrow}$ security
2. Whereas trust is not needed given absolute security, this is difficult to achieve in human-made systems and, when achieved, tends to be over-engineered to the extent that it is not favored in practice. Therefore, trust needs to be introduced as a concept.
3. "Secure" is an objective state ("The web is secure")
   "Trust" is an action ("The web is trusted by …")
   Trust requires someone to do the trusting
4. In the context of data exchange, the sender (S) and receiver (R) are the ones who do the trusting

2. Security and Trust

The terms "security" and "secure" have been widely used in the field of data communications in a sense that is analogous to trust. As shown in **Figure 1**, something that is secure is trusted, but something that is trusted is not necessarily secure. Given complete security, there would be less need to introduce a new concept of trust, at least with regard to specific communications links.

However, an artificial system that never malfunctions is in general very hard to construct in the real world. Even where an infallible theoretical solution does exist, it is often not accepted by society for being non-optimal when economic viability or other human factors are taken into account.

In the case of data communications, quantum cryptography with single photon transmission technologies such as BB84 is assumed as offering the highest level of security. But a global agreement has yet to be made to implement it as the global infrastructure of data communication. When talking about the robustness of widely used public key infrastructure, it is important to refer to human factors such as the leakage of private keys through mistakes made by the holders of keys or through cyber attacks by someone else.

Considering the technical difficulty of building infallible systems and the requirement of society to make an "optimal" solution that takes human factors in account, the authors believe it is natural that attention now be paid to introducing the concept of trust.

3. Modeling DFFT

One more key difference between security and trust is that, whereas describing something as secure indicates an objective state, trust is an action and, as such, requires someone to do the trusting. The authors believe that, in the context of the exchange of data, the "someone" should be the sender or receiver of data.

In many fields such as finance, transactions between enterprises, and the Internet of Things (IoT), it has become common in recent years for the sender and receiver of data to be computers – not humans – but machines in other words. The authors are of the view that, no matter how highly advanced their artificial intelligence is, it should be humans who are responsible for the machines' rules of sending and receiving data, for the program they use to prepare data, and for the program by which that program is generated[4]. Accordingly, for the purposes of this article, the sender and receiver of data refers to humans, doesn't refer to any machines.

There is a debate over whether the data rights holder, the person who collects data from the rights holder, and the person who holds the data copyrights for statistically processing the collected data should be made independent of the sender. However, in cases where another person sends data to a communications channel contrary to the intention of the data rights holder and data copyrights holder, the authors consider the problem where data is passed from another person to the data sender separately from the problem where the data sender sends data to the communications channel. In this case, the model described in this paper applies to data that is distributed via the communications channel. But cases that do not depend on the communications channel are regarded as other human-related problems such as contract problems between two parties.

On the other hand, the object of this trust (what it is that the sender and receiver put their trust in) is assumed to be: (A) the communications channel, (B) the communications partner (the receiver for the sender or the sender for the receiver), and (C) the data being communicated. Narrowly defined, a "communications channel" refers to the communications link itself (ex. electrical, fiber-optical, and wireless) and its directly attached machines (ex. digital-analog converters, analog-digital converters, and protocol converters), it also includes the business entities that operate all of the above. For this article, however, a broader definition of "communications channel" is adopted that includes all the machines in between the human (the sender or receiver) and the above narrowly defined communications channel. They are the machines that generate, send, and receive data in accordance with logic and other rules for which the sender and receiver are responsible (**Figure 2**).

Figure 3 shows the DFFT model devised by the authors. The basic structure of data communications involves a sender of data, a communications channel, and a receiver of data. Note that in the case of many receivers receiving the same data from the same sender, each data communication between a receiver and the sender is considered to be the primitive element where this model holds.

Figure 2 — What is a Communications Channel?

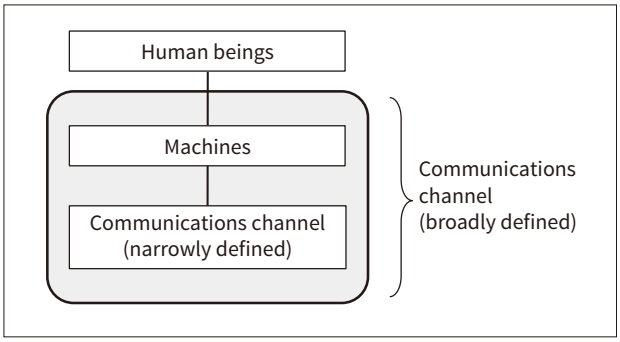
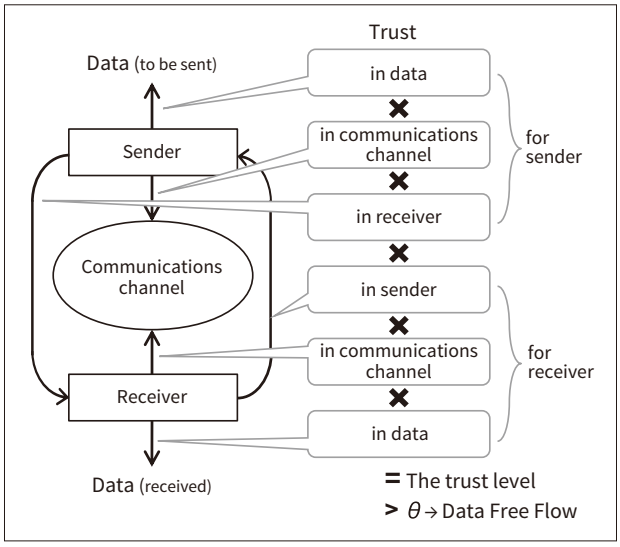


Figure 3 — DFFT Model



As defined above, trust in data communications refers to trust by the sender and receiver in the communications channel, in the communication partner, and in the data being communicated. This can be represented as the six parameters below. They are the extent to which the sender (S) has trust in

- the communications channel: S(A)
- the receiver of the data: S(B)
- the data being communicated: S(C)

and the extent to which the receiver (R) has trust in

- the communications channel: R(A)
- the sender of the data: R(B)
- the data being communicated: R(C).

The "trust level" is then defined as the mathematical product of these six parameters. Accordingly, a free flow of data occurs if this trust level exceeds a certain threshold (or threshold vector), or when humans can be explicitly aware of this level of trust in particular communications scenarios. This is what constitutes our DFFT model.

Regarding quantifying trust, especially in (B) and (C), it is acknowledged that the authors are motivated by a publication issued by the Organisation for Economic Co-operation and Development (OECD) entitled "OECD Guidelines on Measuring Trust"[5].

4. Definition of Trust in Communications Channel

In order to engineer DFFT, the first thing to do is to define S(A) and R(A), the trust in a communications channel by senders and receivers. The "communications channel" refers to the broad definition explained in Chapter 2 and in **Figure 2**. We define S(A) as the extent to which the sender has faith that the system will pass the data to the receiver in the way the sender intends. The authors define R(A) similarly as the extent to which the receiver has faith that the system passed the data the receiver received in the way intended by the sender.

The S(A) is now broken down into the 10 factors shown in **Figure 4 (a)**. In terms of the sender's faith, the security community has long discussed how accurately data is transmitted from sender to receiver [$S(A)_1$], the extent to which data is leaked in transit [$S(A)_2$], and the extent to which it is tampered with in transit [$S(A)_3$]. To these can be added the extent of time delay and the variance of time delays in transmission [$S(A)_4$], which is also a requirement for trust especially in applications where real-time performance is important. The confidence that the person who received the data was in fact the intended recipient [$S(A)_5$] has become an active topic of debate in recent years. The extent to which transmissions can subsequently be rolled back [$S(A)_6$] will be a major issue particularly in corporate information systems to cope with e-mails sent mistakenly by employees.

The extent to which the transmission of data is verified to be legitimate [$S(A)_7$] will be in demand as a means of enhancing trust in communications channels. It is also an important technical issue in knowledge processing shown in the next chapter. The extent to which senders can inadvertently leak the data concerned [$S(A)_8$] is dealt with separately from $S(A)_2$, as it relates to actions taken prior to communications. The extent of negative flow-on effects, unintended by the sender, that occur after the receiver has received the data [$S(A)_9$] is also being considered for addition to the requirements for trust in communications channels. This recognizes the size of the associated social problems, such as the many cases where a message, having been sent to large numbers of recipients, "goes viral" with consequences that can even include people committing suicide. The ability to perform after-the-fact verification of these [$S(A)_{10}$] will also likely be an important future requirement for trust in the flow of data.

The R(A) is broken down into the 10 factors shown in **Figure 4 (b)**. In terms of the extent to which the receiver trusts that the system passed the data they received in the way intended by the sender, how accurately data is transmitted from sender to receiver [$R(A)_1$], the extent to which data is leaked in transit [$R(A)_2$], the extent to which it is tampered with in transit [$R(A)_3$], and the extent of time

Figure 4 — Definition of Trust in Communications Channel

The extent to which the sender has faith that a web system will pass the data to the receiver in the way intended:

$S(A)_1$: How reliably data is passed from S → R
$S(A)_2$: Extent of leakage in transit
$S(A)_3$: Extent of tampering in transit
$S(A)_4$: Confidence that receiver was intended recipient
$S(A)_5$: Time delays from sending by S to receiving by R
$S(A)_6$: Extent to which transmissions can subsequently be rolled back
$S(A)_7$: Ability to verify that the transmission of data was legitimate
$S(A)_8$: Extent to which S can inadvertently leak data
$S(A)_9$: Extent of negative flow-on effects, unintended by S, that occur after R has received the data
$S(A)_{10}$: Ability to perform after-the-fact verification of the above

(a) Sender's trust in communications channel

The extent to which the receiver has faith that a web system has passed the data they received in the way the sender intended:

$R(A)_1$: How reliably data is passed from S → R
$R(A)_2$: Extent of leakage in transit
$R(A)_3$: Extent of tampering in transit
$R(A)_4$: Confidence that sender was who it is believed to be
$R(A)_5$: Time delays from sending by S to receiving by R
$R(A)_6$: Extent to which transmissions can subsequently be rolled back
$R(A)_7$: Ability to verify that the transmission of data was legitimate
$R(A)_8$: Extent to which S can inadvertently leak data
$R(A)_9$: Extent of negative flow-on effects, unintended by R, that occur after R has received the data
$R(A)_{10}$: Ability to perform after-the-fact verification of the above

(b) Receiver's trustin communications channel

delays in transmission from sender to receiver [$R(A)_4$] are the same as the corresponding $S(A)_1$, $S(A)_2$, $S(A)_3$, and $S(A)_4$. The only difference is that they are from the receiver's perspective. Similarly, the confidence that the sender is who it is believed to be [$R(A)_5$], the extent to which transmissions can subsequently be rolled back [$R(A)_6$], the extent to which reception of the data concerned can be verified as legitimate [$R(A)_7$], and the extent to which receivers can inadvertently leak the data concerned [$R(A)_8$] are also equivalent to $S(A)_5$, $S(A)_6$, $S(A)_7$, and $S(A)_8$, only from the opposite perspective.

The extent of unintended flow-on effects after the receiver receives the data [$R(A)_9$] is also a factor that needs to be considered in relation to trust in communications channels in the broad definition. An example of this is how the spread of computer viruses has undermined trust in communications channels. The ability to perform after-the-fact verification of these [$R(A)_{10}$] will, like $S(A)_{10}$, possibly be an important future requirement for trust in the flow of data.

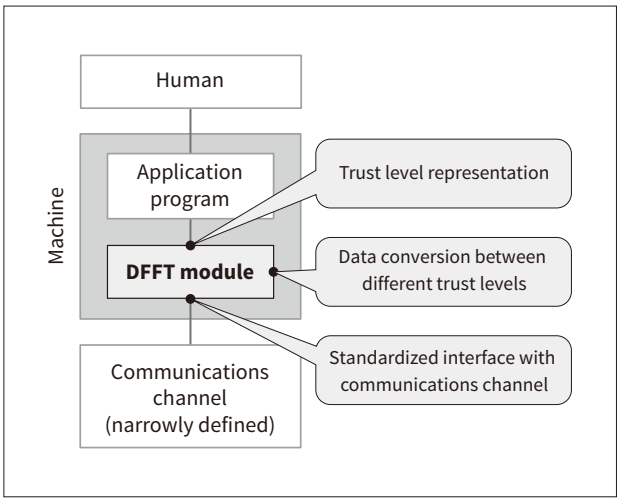
5. Technical Challenges of DFFT and Hitachi's Strategy

The information processing functions of the machine that makes up the broad definition of the communications channel shown in **Figure 2** can be divided into (a) functions that depend on applications and (b) what the authors call the DFFT module, which works independently of applications. The main technical challenges associated with the DFFT module are shown in **Figure 5**.

The first challenge is to design how trust is represented. The trust representation needs to be expanded beyond the unidimensional binary parameters for whether the data is classified or not to encompass multi-dimensional continuous variables. The dimensions will refer to who, where, when, why, and how the data is used in addition to what the data is. Standardization of the access interface to such a trust representation will follow. Given a level of trust, application programs access data through the standardized interface.

The second challenge is to devise the logic to transform the data to suit the required level of trust. It is a generalization of lossy compression when the entropy of the information decreases. For the case where entropy is to be increased, the authors are considering a knowledge system. Such a system will also work to minimize the imposition of extra work on the human senders, receivers, and their application programs when the DFFT module is first implemented.

Figure 5 — Technical Challenges of DFFT



The third challenge is the standardization of the interface between the DFFT module and standardized computer networks. The interface should provide secure communications including human factors and cryptographic mechanisms such as authentication of entities and the integrity/confidentiality of transmitted messages will play a central role. As the authors have explained elsewhere[6] in this edition of *Hitachi Review*, PBI (the Public Biometrics Infrastructure) combining public key cryptography with the state-of-the-art biometrics technologies at the time it is used is a promising candidate.

The research into DFFT described in this article is proceeding under the name of Tokken, which is the Hitachi scheme established 60 years ago to do a highly important R&D project. Current Hitachi R&D is made up of ten technology centers, one fundamental research center, a department to promote external collaborations, and a technology strategy department to manage all the above. Of those ten technology centers, the center for artificial intelligence, the center for communication technologies, and the center for information systems together with the fundamental research center are involved in this Tokken. A supervisory role for the multiple centers is played by a steering team led by the Corporate Chief Engineer and staff from the technology strategy office and the administration office at the central research lab in Tokyo, Japan.

6. Conclusions

This article has described the work on DFFT by Hitachi as one example of the work being done by industry in the private sector. By making this research open, Hitachi hopes to substantially expand collaborative creation with stake holders in industry, government, and academia around the world[7] and to contribute to realizing DFFT.

References

1) "Data Free Flow with Trust (DFFT): Paths towards Free and Trusted Data Flows," The World Economic Forum, Geneva (May 2020), http://www3.weforum.org/docs/WEF_Paths_Towards_Free_and_Trusted_Data%20_Flows_2020.pdf
2) "Updating Governance Mechanism for Trusted Digital Society," The Centre for the Fourth Industrial Revolution Japan, The World Economic Forum, Tokyo (2021).
3) "White Paper," Trusted Web Promotion Council, Cabinet Secretariat (2021) in Japanese.
4) H. Takeda, "Human-oriented Research and Development," Hitachi Review, 58, pp. 126–132 (Sep. 2009).
5) "OECD Guidelines on Measuring Trust," Organisation for Economic Co-operation and Development, Paris (Nov. 2017), https://doi.org/10.1787/9789264278219-en
6) T. Kaji et al., "Trusted and Secure Service System for Society 5.0," Hitachi Review, 70, pp. 457–461 (Jun. 2021).
7) H. Takeda, "Hitachi R&D Strategy," Hitachi Review, 63, pp. 539–547 (Nov. 2014).

Authors

Haruo Takeda
Corporate Chief Engineer, Research & Development Group, Hitachi, Ltd.

Akira Ishikawa
Technology Adviser, Technology Strategy Office, Research & Development Group, Hitachi, Ltd.

Tadashi Kaji
Senior Chief Researcher, Center for Technology Innovation – Systems Engineering, Research & Development Group, Hitachi, Ltd.

Kenta Takahashi
Chief Researcher, Security Research Department, Center for Technology Innovation – Systems Engineering, Research & Development Group, Hitachi, Ltd.

Toshiaki Suzuki
Senior Researcher, Connectivity Research Department, Center for Digital Technology Innovation – Digital Technology, Research & Development Group, Hitachi, Ltd.

Tatsuya Teshima
Senior Researcher, Planning Office, Central Research Laboratory, Research & Development Group, Hitachi, Ltd.

Kyoko Yamamoto
Member, Human Capital Management & General Affairs Division for Research & Development Group, Human Capital Group, Hitachi, Ltd.

Hiromitsu Katou
General Manager, Center for Technology Innovation – Systems Engineering, Research & Development Group, Hitachi, Ltd.

Seishi Hanaoka
General Manager, Center for Digital Technology Innovation – Digital Technology, Research & Development Group, Hitachi, Ltd.

Tatsuhiko Kagehiro
General Manager, Center for Technology Innovation – Artificial Intelligence, Research & Development Group, Hitachi, Ltd.

Shinji Nishimura
General Manager, Center for Exploratory Research, Research & Development Group, Hitachi, Ltd.

Norihiro Suzuki
Vice President and Executive Officer, CTO, and General Manager of Research & Development Group, Hitachi, Ltd.